

La investigación cuantitativa

SERGIO PÉREZ CASTAÑOS Y SANDRA GARCÍA SANTAMARÍA

2.1. Introducción

Tras haber repasado en el capítulo anterior cuáles son las características más relevantes de la investigación cualitativa, le toca el turno a la otra tipología de investigaciones que habitualmente desarrollamos en Ciencias Sociales. Además, este capítulo encaja como paso previo del tercero, el cual versará sobre las metodologías mixtas, en donde se realiza una combinación de estrategias cualitativas y cuantitativas.

Por nuestra parte, en el análisis cuantitativo el objetivo fundamental recae en probar las teorías e hipótesis que desarrollamos examinando relaciones entre variables, habitualmente por medio de procedimientos estadísticos. En términos generales, la aproximación cuantitativa se caracteriza por sostener la existencia de relaciones causales, desarrollando el conocimiento a través de la observación y la medición de la realidad de una manera lo más objetiva u objetivable posible. Tratando de buscar recurrencias que puedan desembocar en generalizaciones que nos ayuden a explicar fenómenos sociales de cualquier índole (López Roldán y Fachelli, 2015).

En este sentido, y siguiendo una estructura muy similar a la existente en el capítulo anterior, este apartado se dividirá en cinco epígrafes, además de esta breve introducción. En el próximo se analizarán cuáles son las características de la metodología cuantitativa a la hora de realizar investigaciones de este tipo. El tercer elemento que destacaremos serán las diferentes fases que

se deben abordar cuando se realiza una investigación de este tipo. Un cuarto apartado, toda vez que ya conoceremos cuáles son las características y fases que seguir, son las técnicas e instrumentos para la recogida de datos. Aunque mencionaremos varias, nos detendremos de forma más profunda en una de las técnicas más habituales en las Ciencias Sociales: la encuesta.

En la medida en que sabremos cómo obtener los mejores datos para nuestro análisis cuantitativo, el apartado número cinco de este capítulo abordará una de las cuestiones fundamentales para la realización de una investigación: la realización de análisis estadísticos. Además de diferentes explicaciones de tipo teórico-matemático, todas las aproximaciones y explicaciones irán conjugadas con demostraciones de *software* de aplicación estadística. Sabemos que existen multitud de *software* de esta tipología, pero en esta cuestión nos centraremos en uno de los más ampliamente utilizados para la docencia e investigación en Ciencias Sociales, que no es otro el programa SPSS¹ de IBM.

Tras conocer los diferentes fundamentos estadísticos, de análisis y de interpretación de datos cuantitativos, el último de los apartados de este segundo capítulo girará en torno a dos conceptos fundamentales, que son clave para poder seguir haciendo ciencia y que nuestras investigaciones cuantitativas puedan ser replicables y útiles para la sociedad: la fiabilidad y la validez. Dentro de este apartado dividiremos no solo entre estas dos cuestiones, sino que hablaremos de estos elementos en la aplicación de las herramientas cuantitativas, por un lado, y en la realización de los análisis, por otro, acompañándolo de apoyo para la realización de diferentes pruebas estadísticas que puedan ayudarnos a la hora de testar nuestras investigaciones.

2.2. Características de la metodología cuantitativa

Las técnicas cuantitativas tienen un encaje fundamental en las investigaciones de las Ciencias Sociales en la medida en que, bus-

1. Más información sobre este paquete estadístico, así como versiones de prueba gratuita para estudiantes, pueden encontrarse en su página web: <https://www.ibm.com/es-es/analytics/spss-statistics-software>

can responder a las preguntas de investigación que se planteen en términos lo más objetivos y racionales posibles. Esta racionalidad se encuentra fundamentada tanto en el cientificismo y el racionalismo como en posturas epistemológicas institucionalistas (Asimov, 1979; Beltrán, 2000). En este sentido, se abrazan numerosos elementos vinculados a las Ciencias exactas, como es el uso de la neutralidad valorativa como criterio de objetividad, debido a que el conocimiento se fundamenta en hechos, prescindiendo de la subjetividad de los individuos.

A pesar de que este enfoque tiende a ser criticado, porque realiza una representación de la realidad parcial y atomizada, arrojando al experto el halo de autoridad, no es menos cierto que sus características dibujan esta aproximación con un carácter lineal y centrado en evidencias. Así, cumpliendo con las características que se expondrán a continuación y siguiendo de forma correcta las fases establecidas en el tercero de los epígrafes, podemos establecer que existirá una claridad expositiva entre los elementos que conforman el problema de investigación, la definición de estos, sus limitaciones y saber con exactitud dónde se inicia el problema, así como el tipo de incidencia existe entre sus elementos (Price y Murnan, 2004).

Para poder alcanzar estos objetivos genéricos, el análisis cuantitativo necesita examinar los datos de manera numérica y, para ello, debe apoyarse en la estadística. Utilizaremos, entonces, elementos fundamentales para la investigación como son las variables, la correlación y las unidades de observación. ¿Y por qué la estadística? Porque solo a partir de ella podremos realizar inferencias gracias a la transformación de elementos teóricos en valores numéricos (Castro Posada, 2002). De esta manera aportaremos objetividad a las investigaciones al realizar una medición exhaustiva y controlada, intentando buscar la certeza en el conocimiento. Del mismo modo, gracias a la posición externa de las personas que realizan la investigación, existe una relación de independencia entre sujeto y objeto. Es importante recalcar la importancia de tratar de alejarse lo máximo posible del objeto, porque, de lo contrario, estaríamos realizando una observación participante.

No obstante, no deben abandonarse los preceptos teóricos, dado que la teoría es el elemento fundamental de la investigación social (Purtois y Desnet, 1992). En la teoría encontramos el

origen y nos ayuda a definir los elementos clave de nuestra investigación; es decir, nos aporta el marco que va a encuadrar nuestro análisis. Asimismo, el fin último de las investigaciones es convertirse en eso mismo, en teoría que aporte y abunde de cara a la realización de futuras investigaciones, así como de cara a la resolución de diferentes problemas sociales. Gracias a un dominio adecuado de la teoría podremos aportar una comprensión explicativa y predicativa de la realidad.

De esta forma, una de las características generales más importantes es que el análisis cuantitativo utiliza un método hipotético-deductivo. Esto quiere decir que, a pesar de que la ciencia se basa en la inducción sistemática, es también, en gran medida, deductiva. Es decir, el relacionar datos, establecer conceptos y enunciados a partir de ellos y sacar conclusiones de todo género es un trabajo deductivo. Esto es, de problemas planteados como hipótesis (elementos que establecen la relación de unas cuestiones sobre otras) se infieren conclusiones que acabarán revirtiendo en la sociedad y en futuras investigaciones.

Habiendo clarificado las características más generales de las metodologías cuantitativas, es ahora el momento de conocer los componentes y características concretas más comunes en este tipo de aproximaciones. Es fundamental entender que no existe una homogeneidad interna y, por lo tanto, no todas se aplican en todos los momentos (Rossi, 1982). Sí que es cierto que los que enunciaremos a continuación son elementos compartidos por todas las técnicas y herramientas que pueden ser englobadas en esta tipología de análisis.

La primera de las características comunes es que este tipo de análisis centran su atención en la representatividad del dato, como eje para poder efectuar generalizaciones. Solo a través del dato es posible cumplir con los requisitos estadísticos para poder realizar inferencias y, por ende, generalizaciones sobre el total de la sociedad (Giner, 1983). De esta manera, se da la segunda de las características, estrechamente imbricada con la anterior. Esta no es otra que el hecho de que esta aproximación asume que los fenómenos a analizar se dan de forma regular, o sea, se producen de forma cíclica y similar en todos los lugares (Cicourel, 2011).

Esta regularidad en la aparición de los hechos lleva a la tercera característica, que versa sobre el valor relativo que adquieren los

contextos naturales a la hora de explicar fenómenos (Lazarsfeld y Oberschall, 1965). En todo caso, cada vez en más y más disciplinas de las Ciencias Sociales nos encontramos con investigaciones que incluyen los factores contextuales en su análisis. Esto se debe a que, a pesar de que exista determinada periodicidad en los fenómenos a analizar, puedan darse características especiales o excepcionales en función del medio en el que se desarrollan.²

Encadenado con la segunda de las características se encuentra la cuarta, en la medida en que se centra el interés en aquellos elementos que pasan por ser repetitivos, frecuentes y normales (esto es, que se den de forma similar o constante). Con ello se pretende poder ganar en capacidad explicativa y que investigaciones puedan ser extrapolables para diferentes casos de estudio (Bonilla Castro y Rodríguez Sehk, 1997). Así se genera el caldo de cultivo para la quinta de las características, dado que el análisis cuantitativo pone el acento en la predicción de fenómenos y la capacidad explicativa de sus análisis ante lo que se pretende analizar (López-Roldán y Fachelli, 2015), precisamente gracias a esa posibilidad de frecuencia y normalidad.

La sexta y séptima características específicas van de la mano en lo que a la utilización de la población se refieren. La sexta establece que cuando realizamos investigaciones cuantitativas, tendemos a enfatizar las acciones que llevamos a cabo en aras de identificar las formas de distribución de la sociedad (Cicourel, 2011). Gracias a poder realizar estas identificaciones se puede dar paso a la séptima, en la medida en que, tomando la clasificación, anotación, medición y conteo como operaciones fundamentales, podremos realizar comparaciones entre poblaciones o entre grupos conformantes de dichas poblaciones (Giner, 1983).

El último bloque de elementos comunes a la investigación cuantitativa engloba cuatro características fundamentales. La primera gira en torno a que, las investigaciones cuantitativas ponen su foco más en los resultados que en los procesos; es decir, que la obtención de determinados resultados es lo más relevante para el análisis, más allá de todo lo que pueda rodear a la investigación. Además, las personas que se encuentran realizando este tipo de estudios tienden a definir las cuestiones metodológicas a

2. El contexto no solo se refiere al plano físico-geográfico, sino también a aquello que rodea al objeto de estudio tanto física como psicológicamente.

priori; es decir, una vez que se topan con un problema social, buscan la mejor herramienta basada en la metodología para poder obtener, procesar y analizar la información (Weber, 2012). De esta manera, se da paso a la tercera de las características, que no es otra que, se parte de una serie de hipótesis que se plantean en las primeras fases del proceso de investigación (que desarrollaremos en el siguiente apartado) y tratan de validarlas a través de los datos. Finalmente, todo esto es posible porque, cuando se plantean este tipo de investigaciones, se hace desde una perspectiva externa, sin ánimo de alterar el contexto que se analiza. Esto se consigue gracias a que la toma de datos suele ser lo menos invasiva posible y, posteriormente, estos datos son analizados a través de un procedimiento experimental que busca poder objetivar lo máximo posible las conclusiones que se extraigan del análisis.

Todas estas características acaban por estar incluidas en los diferentes procedimientos y operaciones que se van secuenciando a lo largo del desarrollo de la investigación y su metodología (Crespo, 2016). Así, conforme se van dando los diferentes pasos que se analizarán en el próximo apartado, se acabará por definir el caso de estudio, la población objeto de análisis, la herramienta más útil para su realización y, finalmente, se adoptarán las técnicas más adecuadas para poder realizar un análisis adecuado del problema que se va a analizar.

Dos elementos fundamentales para poder realizar un análisis cuantitativo son conocer cuáles son nuestra población o universo y decidir cuál será nuestra muestra. Habitualmente, se define *universo* como el conjunto de individuos u objetos de los cuales se desea conocer algo (Camel 1970), si bien las definiciones también suelen hacer hincapié en las características de los objetos, susceptibles de ser estudiadas (Polit y Hungler 1987). Así, el universo puede ser finito o infinito. En el primero de los casos, su estudio podrá hacerse a través de censos o muestras; cuando es infinito, solo por el segundo método. Los conceptos de *finito* e *infinito* en la metodología poseen una carga relativa, pues en algunos casos no es posible definir plenamente los límites del universo. Para ciertos tipos de estudio, cuando el universo no es excesivamente extenso, el censo no solo es posible, sino a veces aconsejable, pero tiene la desventaja de los costes y tiempos implicados.

De nuestro universo se extraen las muestras, de las cuales habrán de inferirse los resultados. Una buena muestra es aquella que es representativa de la población objeto de estudio (del universo); no obstante, los criterios de representatividad no son unívocos. Generalmente las metodologías cuantitativas se basan en criterios sustentados en la diversidad del universo, es decir, una buena muestra es aquella que puede dar cuenta de toda la diversidad existente, por lo tanto, cuanto más diverso sea el universo, mayor (tanto en diversidad como en tamaño) será la muestra. De la misma manera, para poder cubrir esa diversidad existente, el nivel de error que se esté dispuesto a aceptar debe ser el menor posible, por lo que mayor deberá ser el tamaño muestral.

Sin embargo, no todo son características positivas para este tipo de investigaciones y también existen diferentes límites que deberíamos tomar en consideración a la hora de plantear este tipo de análisis. Para Campbell y Stanley (2002), existen al menos tres limitaciones de la investigación cuantitativa que se dan, de forma generaliza en cualquier proceso de estas características.

La primera es el hecho de que se dan investigaciones cuantitativas que tratan de disfrazar la subjetividad individual de las personas que se encuentran realizando la investigación a través del traje de la científicidad. Esto es, que se trata de utilizar este tipo de métodos de análisis para tratar de sesgar los resultados obtenidos o tratar de obtener conclusiones preconcebidas ante la problemática que se pretende analizar (por ej.: dar una pátina de credibilidad a prejuicios sociales).

Otra de las limitaciones que los autores encuentran en este tipo de análisis es la conjugación cuantitativa de agrupaciones para estudiar los sistemas sociales. Esto es, al tomar a las personas como objeto, las diferencias individuales y culturales entre grupos no pueden promediarse estadísticamente (vinculado, entre otras cuestiones, a la toma en consideración del contexto). La tercera de las limitaciones es la de tomar una parte del sistema como el todo, de tal manera que se obvia la naturaleza fundamental de los fenómenos sociales, que no es otra que la de ser compleja y cambiante. Así, en ocasiones se puede pretender generalizar unos resultados que pueden ser excepcionales a partir de condicionantes preexistentes.

2.3. Fases procedimentales de la investigación cuantitativa

Ya sabemos cuáles son las características generales que debe tener una investigación cuantitativa y conocemos el papel central de los datos y de la utilización de la estadística, en los que profundizaremos en los epígrafes cuatro y cinco de este capítulo. Ahora es momento de conocer cuáles son las diferentes fases que enfrenta una persona cuando realiza una investigación cuantitativa. La investigación requiere tiempo, recursos humanos y materiales y, sobre todo, conocer los pasos que se deben ir dando para poder tener éxito en ella.

En términos generales, y con elementos compartidos con lo desarrollado en el anterior capítulo, la investigación es un proceso destinado a resolver inquietudes a partir de la recolección y análisis de información. Estas inquietudes, muchas veces convertidas en dilemas sociales o problemáticas que pueden poner en jaque nuestro sistema social, tienen en la investigación una fórmula para ser descritas, interpretadas y predichas. Así, la investigación en general produce conocimiento; sin embargo, aquellas que se realizan bajo un prisma cuantitativo obtienen este conocimiento basándose en datos, lo que nos permite tomar decisiones más razonables, averiguar qué sucede más allá de nuestros ojos y tratar de predecir qué sucederá en el futuro (Crespo, 2016).

Antes de comenzar a dar pasos en nuestra investigación debemos centrarla, y es que una investigación en Ciencias Sociales puede girar en torno a cualquier ámbito de la sociedad. Las preguntas que nos podemos realizar y que podemos buscar responder son innumerables y abarcan un sinnúmero de campos temáticos que van desde por qué la ciudadanía decide tomar una decisión o votar a una determinada fuerza política hasta las intrincaciones teóricas de por qué el ser humano decide vivir grupalmente o por qué las sociedades se organizan de esta o de aquella manera (Martínez, 2016).

El único punto del que todos estos temas pueden ser descritos es la ciencia y el método con el que realizar la investigación, el método científico. La ciencia nos lleva a empujar nuestras barreras del conocimiento, buscando dar respuesta a nuevos enigmas, y posee una serie de objetivos. Una vez que los hemos definido,

debemos responder a las preguntas sobre qué tipo de investigación será la nuestra. Así, las cuatro preguntas a las que se suele dar respuesta son: ¿qué es?, ¿cómo es?, ¿por qué sucede? y ¿cómo debe ser? Algunas de las investigaciones únicamente darán respuesta a alguna de estas preguntas, mientras que otras pueden buscar responderlas todas (Batthyány y Cabrera, 2011).

La selección del problema de investigación es siempre una cuestión complicada a la que se enfrenta cualquier persona desarrollando una investigación, ya sea novel o experimentada. Es importante exponerlo de forma clara, dado que facilitará la investigación. ¿Por qué? Muy sencillo, puesto que en la medida en que los diferentes elementos queden claramente establecidos, la búsqueda de la información, los datos y el desarrollo de la investigación serán lo más sencillo posible.

Por ello hay dos cuestiones fundamentales que deben ser desarrolladas. Por un lado, tenemos el tema de estudio o análisis, que sería el ámbito de la rama de las Ciencias Sociales en la que estamos realizando nuestra investigación; esto es, en donde se sitúa el problema (comportamiento electoral, desviaciones sociales, movimientos sociales, construcciones familiares, etc.). Por el otro, nos encontramos con el caso, que sería la parte específica de la realidad en la que se centra nuestra investigación (Europa, España, colegios públicos...)³ (Manheim y Rich, 2002).

Pero, además de estos dos elementos, hace falta algo más, y es que solo con esto no tenemos nada a qué responder. Únicamente sabemos dónde y cómo vamos a acercarnos al problema, pero ¿qué queremos responder? Aquí es donde surge la pregunta de investigación. Nuestra pregunta debe tener una serie de características (Anduiza, Crespo y Méndez, 2011):

- No puede ser ni demasiado amplia ni demasiado concreta. Clara y acotada.
- Susceptible de tener una respuesta.
- Hay que justificar la relevancia del tema (importancia sustantiva y valor teórico). La justificación personal no es suficiente, aunque sea la motivación principal de quien investiga. La justificación tiene que residir en la importancia del problema:

3. No tiene por qué ser una unidad geográfica, puede ser cualquier elemento o agrupación de elementos.

¿por qué es relevante resolver esa pregunta?, ¿qué implicaciones se derivan del estudio?

- Debe ser original, es decir, si replica algo que ya se ha hecho sobre el mismo caso y se llega a idénticas conclusiones, no tiene valor.
- No hay que incluir la referencia al caso de estudio en el problema de investigación.

Una vez definido el problema aún nos enfrentamos a cinco nuevas etapas que detallaremos a continuación. Así, para conseguir llegar a resultados que sean aplicables a nuestra sociedad, toda investigación, sea esta del tipo que sea, sigue un proceso circular.⁴ Esto quiere decir que mantiene una relación constante entre experiencia y teoría. La teoría se prueba mediante datos empíricos y estos se interpretan sobre la base de aquella. Una forma visual de entenderlo es a través del siguiente diagrama representado en la figura 1.

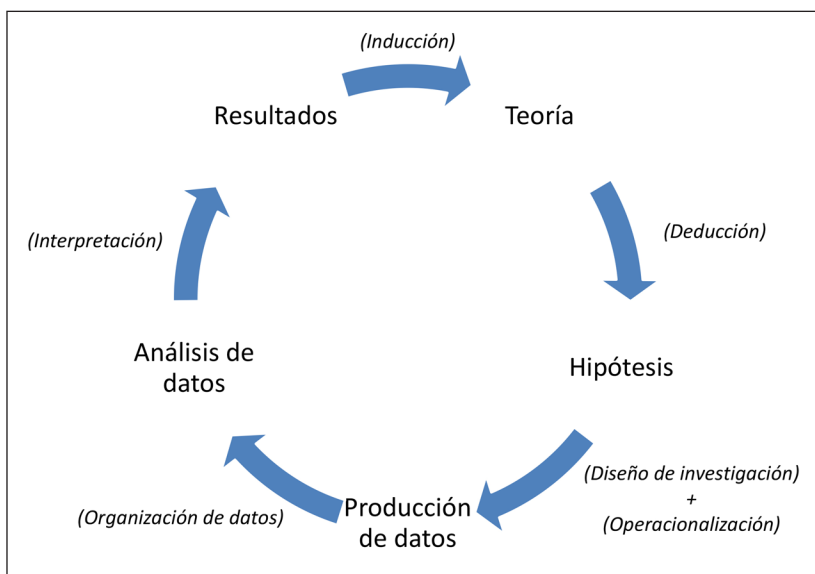


Figura 1. Circularidad del proceso de investigación. Fuente: elaboración propia.

4. La característica de ser circular es una de las asociadas al método científico que, si bien no ha sido tratado en este capítulo, se debe aplicar en cualquier proceso científico, ya sea de Ciencias Exactas o de Ciencias Sociales.

Según Corbetta (2007), la investigación científica es un proceso de descubrimiento creativo que sigue un itinerario prefijado y unos procedimientos preestablecidos y consolidados dentro de la comunidad científica. Por ello, descubrir problemas y teorías es un proceso creativo, no sujeto a reglas; sin embargo, comprobarlas sí que sigue unas normas precisas: hay una estructura lógica en el proceso de investigación y unas técnicas específicas. Así, existen cinco fases diferenciadas a través de las cuales vamos dando pasos para la construcción de nuestra teoría, su comprobación y su asentamiento o no.

En la medida en que, como hemos establecido en el anterior apartado, las investigaciones cuantitativas tienen una aproximación hipotético-deductiva, tomaremos nuestro punto de partida en la deducción; esto es, en el paso de la teoría a las hipótesis. Pero ¿qué es la teoría? Es un conjunto de proposiciones relacionadas que se formulan con un alto grado de abstracción y generalización respecto a la realidad, que proceden de hechos empíricos que se producen con regularidad y de las que pueden inferirse predicciones empíricas (Manheim y Rich, 2002). Es decir, contamos con afirmaciones con una estructura causal que van más allá de expresiones empíricas concretas que surgen de la constatación de repeticiones de la realidad social. Además, gracias a la teoría podemos deducir hechos en otros contextos diferentes.

Para poder desarrollar las diferentes fases de una investigación de forma adecuada, es imprescindible analizar la literatura existente en relación con el tema objeto de estudio. Habrá que conocer los libros, artículos, así como las investigaciones e informes que pueda haber al respecto. Más allá de aumentar nuestro conocimiento, el análisis de la literatura existente sobre el tema nos permitirá conocer los aspectos o fenómenos que hayan sido estudiados con anterioridad y nos facilitará concretar de forma exacta el objeto de la investigación (Hernández Sampieri, Fernández-Collado y Baptista Lucio, 2010). Con toda esta información, se acabará por formular mejores y más completas hipótesis. Igualmente, y de cara a las fases finales de la investigación, la literatura nos ayudará a interpretar de manera fidedigna los datos, así como las posibles relaciones entre los conceptos o variables objeto de estudio (Castro Posada, 2001).

El análisis de la literatura existente nos puede ofrecer información válida respecto del mejor diseño, estrategia de medición

y técnicas a utilizar para el análisis que pretendemos realizar. Aunque los estudios precedentes pueden haber cometido errores que normalmente no se explicitan, la lectura de estas nos puede ayudar a mejorar los resultados de nuestras investigaciones (Pignataro, 2016). Del mismo modo, una buena interpretación de la teoría nos puede ayudar a evitar el gasto innecesario de tiempo y dinero, ya que muchos de los datos, análisis e interpretaciones que necesitamos en nuestra investigación pueden estar a nuestro alcance sin necesidad de que tengamos que recopilarlos nuevamente.

De esta manera, partiendo de la literatura y través del proceso deductivo, llegamos al segundo de los pasos: la formulación de hipótesis. Una hipótesis es esencialmente un enunciado de lo que, según creemos, corresponde a los hechos. Expresa lo que esperamos averiguar cuando efectuamos observaciones de la realidad adecuadamente organizadas. Así, las hipótesis son frases declarativas que indican relaciones previstas entre los fenómenos a los que se refieren nuestros conceptos. Por lo general se expresan de la forma siguiente:

«Cuanto más alto –más bajo, mayor, menos, más amplio, más rápido etc.– es -----, más alto –más bajo, mayor, más amplio, más lento etc.– es -----».

Las hipótesis en la medida en que son afirmaciones que proponemos antes de realizar la investigación, pueden ser correctas o incorrectas. En este sentido, será en la fase de interpretación de resultados donde se aceptarán o se rechazarán las hipótesis propuestas (Frías-Navarro, 2008).

A pesar de que, en términos generales, las investigaciones cuantitativas tienden a partir del método deductivo, debemos aclarar que podemos llegar a nuestras hipótesis bien por el razonamiento inductivo o el deductivo. Si empleamos el método del ensayo y el error para construir una teoría, podríamos elaborar la hipótesis mediante un proceso de generalización inductiva: por ejemplo, si observamos que la participación política de la población vasca varía dependiendo de la importancia del primer sector en la economía del municipio, podremos establecer que una mayor o menor importancia del primer sector influye en la participación política. La inducción nos sirve para explicar de

forma indiciaria la relación entre dos variables (Corbetta, 2007). A partir del ejemplo anterior, más encuadrado en el escenario de la Ciencia Política, podemos decir que, hipotéticamente, el tipo de economía del municipio es una de las variables que explican las diferencias entre municipios, pero no nos muestra cómo, o sea, por qué sucede.

Una vez que hemos establecido una teoría que ponga en relación las variables, podemos deducir hipótesis de dicha relación mediante el razonamiento deductivo. El razonamiento deductivo nos dará las hipótesis del por qué (Emmerich, 1997). Utilizaremos otros estudios y otros conocimientos para realizar la deducción. Por ejemplo, si está comprobado la participación política viene determinada por el interés que la ciudadanía tiene ante las cuestiones, noticias y debates políticos, y observamos que en los municipios más rurales los niveles de interés por la política son muy bajos, podremos decir que la participación en los municipios más rurales es más baja, porque existe un bajo interés por la política. En este caso la relación que encontramos ya no es indiciaria, si no basada en la teoría y, por lo tanto, la construcción de nuestras hipótesis ha seguido un procedimiento deductivo.

En general, podemos hablar de cuatro tipos de hipótesis: las *hipótesis nulas*, las *correlativas*, las *direccionales* y las *causales* (Manheim y Rich, 2002). Las hipótesis nulas son el tipo de hipótesis que afirma que no hay relación entre las variables o conceptos analizados. Por su parte, las correlativas son aquellas que sostienen que existe relación entre dos o más variables, pero que no dicen nada respecto a la naturaleza de esta. Una hipótesis correlativa puede ser, por ejemplo, que exista relación entre la edad y la escritura, pero no nos señala si son los jóvenes o los más mayores los que escriben menos o más.

La tercera de las hipótesis son las direccionales. En este tipo de hipótesis las personas que diseñan la investigación señalan la dirección de la relación entre las variables analizadas. Si se espera que ambos elementos aumentan o decrezcan conjuntamente, nos encontramos con una hipótesis de relación directa (cuanto menor es el número de personas analfabetas de una población, mayor es el nivel de desarrollo humano de su país); por el contrario, si esperamos que, a medida que una variable aumenta, la otra decrezca, nos hallamos ante una hipótesis de relación inver-

sa (según aumenta la edad, la capacidad imaginativa de las personas va en detrimento).

Finalmente, la cuarta tipología de hipótesis que podemos plantear son las denominadas *causales*. Estas son el tipo de hipótesis que establecen la relación más atrevida entre dos variables. Es, asimismo, la más difícil de confirmar. Las hipótesis causales pueden tener diferentes formas, en la medida en que pueden, por una parte, señalar simplemente que una variable es causa de otra («La alfabetización crea movimientos democráticos en un país»). En segundo lugar, pueden formularse en términos de «si... entonces...» («Si la renta per cápita de un país es de más de 3000 \$, el número de ONG es superior a 200»). Finalmente, una hipótesis causal puede formularse negativamente («Si en las elecciones generales el Partido Socialista no saca una ventaja de más de 10 puntos al siguiente partido más votado, pierde, por lo menos, un 10% de votos en las siguientes elecciones locales»).

Para poder generar nuestros datos o acudir a otro tipo de fuentes que se detallarán en el próximo apartado, tenemos que convertir nuestras hipótesis en elementos medibles. Para ello, la siguiente fase de la investigación (y una de las más relevantes) es la operacionalización de las variables (Moreno y Mora, 2016). Aquí traducimos nuestras variables de la hipótesis en indicadores que nos sirvan para comprobarlas. Se trata de, partiendo del nivel abstracto de las preguntas, realizar algunas observaciones concretas que nos permitan responder a ellas. Este proceso debe realizarse con suma atención, dado que la concreción en elementos medibles de la variable debe ser correcta; es decir, las variables y los indicadores han de ajustarse a aquello que hemos enunciado en la hipótesis, en la medida de lo posible.

Las hipótesis se construyen a través de conceptos, que son términos teóricos que tenemos en nuestra revisión bibliográfica realizada. Así, cuando queramos comprobar cuál es la traslación de la teoría a hipótesis, utilizaremos ciertos conceptos para poder trasladar de la teoría a la práctica. Por ejemplo, cuando tratamos de comprobar la profesionalización de los políticos, el término *profesionalización* es abstracto y no se puede cuantificar. No tenemos una profesionalización, dos, tres... Hemos de buscar de qué manera el concepto *profesionalización* puede ser medido.

Ahí es cuando damos el salto de la teoría a la realidad, pasando del concepto (en este caso, la profesionalización) a la varia-

ble (Osgood, Suci y Tannenbaum, 1957). Ella nos explica cómo vamos a cuantificar nuestro concepto. Para nuestro ejemplo, una variable que puede medir la profesionalización puede ser el tiempo que se pasa detentando un cargo público. Esto nos permite observar empíricamente el concepto de *profesionalización*.

El último paso es convertir la variable que representa nuestro concepto en unidades de medida cuantificables empíricamente observables. Así, para nuestra variable, tiempo en el que se detenta un cargo público, el indicador será número de años en el cargo público. Así, el indicador es la traslación al campo cuantitativo o empírico de la teoría, pudiendo proceder a la observación y a la comprobación de nuestras hipótesis.

Cuando hablamos de *observación* en una investigación nos referimos al proceso de aplicar un instrumento de medida para asignar valores de alguna característica o propiedad del fenómeno en cuestión a los casos estudiados (Robles Sastre, 2002). O sea, no hablamos solo de la observación como una técnica específica, sino de la aplicación de las técnicas, es decir, de los instrumentos para comprobar que esas proposiciones teóricas son correctas o incorrectas. Todo este proceso es anterior a la aplicación de ninguna técnica.

La operacionalización, dependiendo de la técnica (instrumento) que estemos utilizando tiene que ver con diferentes elementos. En el caso de la encuesta, la operacionalización de las variables tiene que ver con la formulación de las preguntas. En la entrevista en profundidad y el grupo de discusión, esta fase hace referencia a la construcción del guion a utilizar. En la entrevista por otro lado, esta fase se identificará con el protocolo de recogida de datos.

Este proceso debe realizarse con suma atención, dado que la concreción en elementos mesurables de la variable ha de ser correcta; es decir, las variables y los indicadores deben, en la medida de lo posible, ajustarse a aquello que hemos enunciado en la hipótesis. La operacionalización es una fase crucial, dado que la observación se organizará en consonancia con las variables e indicadores planteados (Moreno y Mora, 2016). Así, los errores más habituales que se pueden producir cuando se realiza una operacionalización son que el paso de concepto a la variable no tenga exactitud (o sea, que el paso de lo teórico o abstracto a lo tangible sea inexacto, que el paso de la variable al indicador sea

inexacto), que, a pesar de que hayamos definido bien lo que queremos medir, no utilicemos los elementos correctos para ello, y que los indicadores escogidos no reflejan totalmente el concepto que debemos medir. Debemos tener cuidado en el proceso de operacionalización, dado que la pérdida de significado es muy habitual, se debe tratar de no perder la esencia del concepto.

Todos estos desajustes pueden echar por tierra una investigación. Es uno de los momentos clave del proceso de investigación. Para estar seguros de que no cometemos ningún error y garantizar que nuestra investigación comience con buen pie, tenemos que estar en condiciones de decir a los demás exactamente lo que hemos hecho para obtener nuestras medidas. Si tenemos ayudantes que obtengan la información, las instrucciones que les demos habrán de ser lo más detalladas y precisas posibles como para que cada uno de ellos realice la medición exactamente del mismo modo que los demás. Solo así evitaremos una pérdida de significado.

Por último, una indicación, precisa y pormenorizada del modo de operacionalizar las variables nos ayuda a evaluar los resultados que obtengamos y a eliminar las posibilidades de que los hallazgos se produzcan por defectos del proceso de medición, esto es, tendrán validez.

Si hemos realizado todos estos pasos de forma correcta, podremos avanzar hacia la siguiente fase en la investigación cuantitativa que, tal y como se apreciaba en el diagrama representado en la figura 1, no es otra que la producción de datos, su ordenación y su análisis. Una vez que ya sabemos qué medir, se recolectan los datos, los cuales pueden ser obtenidos a partir de fuentes primarias (nosotros los generamos) o fuentes secundarias (los datos son generados por otras investigaciones o instituciones). Los datos recolectados son procesados y guardados en hojas de variables y datos (Samaja, 2003). Para procesar, guardar y posteriormente analizar los datos se usan programas informáticos como Excel, SPSS, Stata, PSPP o R, entre otros. El análisis de los datos se desarrolla según el tipo de variables y el número de variables a analizar. Las técnicas estadísticas para analizar una variable son tablas de frecuencias, media y desviación estándar, y para analizar dos variables son tablas de contingencia, correlaciones y regresiones, entre otras (Lynch, 2013).

Con el análisis de los datos se tiene como meta responder las preguntas de investigación e identificar tendencias. Una vez analizados los datos, deben presentarse los resultados de forma escrita en un informe de investigación, y con ello culminamos el proceso de investigación. Todo esto será tratado en los siguientes epígrafes de nuestro capítulo y, por lo tanto, no vamos a ahondar en ello ahora. Sin embargo, en aras de comprender el encaje de las investigaciones cuantitativas con los diferentes métodos y aproximaciones que nos permiten las Ciencias Sociales, debemos remarcar el papel que posee el método estadístico (Wonnacott y Wonnacott, 2004).

Este método se basa en la utilización de técnicas matemáticas para descubrir las relaciones existentes entre las diferentes variables. Para poder realizar estas comprobaciones, se manipulan las variables objeto de estudio; si bien debe aclararse que esta manipulación se realiza únicamente de manera matemática. Así, para comprobar la causalidad y los efectos de las variables independientes, el proceso es la subdivisión de la muestra en tantas submuestras como categorías tenga la variable de control, estimándose la relación en cada una de ellas entre las variables dependientes y las independientes (Carrasco Arroyo, 2012).

Para poder realizar todas estas estimaciones, resulta necesario disponer de un elevado número de casos u observaciones, lo que hace que sea uno de los mayores condicionantes a la hora de implementarlo en análisis sociales. Y es que, si nuestra unidad de análisis es diferente a los individuos, hemos de lograr suficiente información al respecto para implementarlo (Bourque y Clark, 2008).

A través de este método, no solo se mide el efecto de las diferentes variables independientes respecto de la dependiente, sino que también nos sirve para conocer si entre ellas matizan sus efectos. Por ejemplo, parece claro que, si queremos ver la influencia del nivel de estudios y la edad en el nivel de ingresos, conforme más edad se tenga, mayor será el nivel de ingresos; pero esa información se puede ver matizada al comprobar la influencia del nivel educacional, en la medida en que una mayor educación también desembocaría en mayores ingresos y, los individuos analizados, en función de sus condiciones etarias, verán cumplimentado un nivel educativo u otro, arrojando ciertas diferencias.

A través de esta subdivisión, se utilizan las probabilidades de que las variables independientes yerren al rechazar la hipótesis nula⁵ de nuestra investigación. De este modo, a través de la prueba de cada una de ellas con las variables propuestas, se comprueba si la hipótesis planteada es correcta o no. Una ausencia de relación entre las variables independientes con la dependiente generará un rechazo de nuestra hipótesis, mientras que, si se descarta en más momentos la hipótesis nula, nos encontraremos con la comprobación de nuestra hipótesis.

2.4. Técnicas e instrumentos de recogida de datos

A pesar de que existen multitud de técnicas de recogida de datos, habitualmente en las Ciencias Sociales se trabaja con dos tipologías de fuentes de datos. Los datos son los elementos de información que obtenemos de cada caso fruto de la observación y medición de nuestro caso de estudio. Podemos distinguir entre *datos primarios* y *datos secundarios* (Emmerich, 1997). Los primarios hacen referencia a los recogidos por las personas encargadas de la investigación, si bien también pueden haber sido recogidos de primera mano en otras investigaciones y haber sido utilizados en ellas, y que, al utilizarlos nosotros, también sean considerados fuentes primarias (por ej.: aquellas encuestas que, pese a haber sido obtenidas por un equipo de investigación y explotadas por este, se utilizan para la realización de una investigación diferente a las realizadas). Por su parte, los datos secundarios hacen referencia a aquellos que han sido obtenidos de otras fuentes tales como archivos, registros, estadísticas, encuestas oficiales o investigaciones en curso. Es probable que este tipo de datos se encuentren ya tratados y codificados de una manera previa, por lo que quien realiza la investigación debe adecuar la informa-

5. Como se ha indicado con anterioridad, la hipótesis nula es aquella que establece la no existencia de relación entre nuestras variables. En términos de contraste y, como explicaremos en el apartado de este capítulo destinado al análisis e interpretación de resultados, siempre hablaremos de aceptar o rechazar la hipótesis nula cuando queramos saber si nuestros planteamientos se cumplen. Aceptaremos la hipótesis nula cuando nuestras variables no tengan relación estadística significativa y la rechazaremos cuando sí que la tengan. Sin embargo, rechazar la hipótesis nula no implica *per se* que nuestra hipótesis original –o alternativa– haya quedado probada.

ción de la que se dispone a las necesidades de la investigación que va a desarrollar. En este sentido, algunas de las técnicas que aquí mencionaremos se desempeñan en ambas tipologías de fuentes; de hecho, es muy habitual que lo hagan en las dos categorías en la medida en que las fuentes primarias de determinadas investigaciones acaben por convertirse en las secundarias de otras (Manheim y Rich, 2002).

La investigación no tiene sentido sin las técnicas de recolección de datos. Estas conducen a la verificación del problema planteado. Cada tipo de investigación determinará las técnicas a utilizar y cada técnica establece sus herramientas, instrumentos o medios que serán empleados (Murillo *et al.*, 2014). En este sentido, y antes de profundizar un poco más en las técnicas más habituales, cabe destacar que estas serán válidas cuando midan lo que realmente desean medir,⁶ es donde redundará su eficacia para predecir el comportamiento de los fenómenos que estudiamos y serán confiables si, al volverse a aplicar, el resultado debería ser muy parecido o similar (Fernández Esquinas, 2003).

En términos generales, un instrumento de recolección de datos es, en principio, cualquier recurso de que pueda valerse quien realiza la investigación para acercarse a los fenómenos y extraer de ellos información. De una forma más simplificada, podríamos definir los instrumentos como los medios materiales que se emplean para recoger y almacenar la información; por ejemplo, fichas, formatos de cuestionario, guías de entrevista, escalas de actitudes u opiniones, etc. (Manheim y Rich, 2002). Entre las fuentes de datos más comunes, encontramos las siguientes:

- *Censos y registros*: esta información, generalmente relativa a la población, las viviendas o la información de los edificios, tiende a encontrarse en prácticamente todos los Estados. El gran inconveniente de los datos censales es que, cuando se encuentran accesibles, se suelen hallar anonimizados, por lo que se pierde gran cantidad de información. Entre sus ventajas destacan la fiabilidad de la información que arrojan, la posibilidad de acceso a tendencias históricas sobre los mismos datos y la comparación entre datos intracensales y con

6. En el epígrafe 6 profundizaremos más sobre los conceptos de *fiabilidad* y *validez* a la hora de realizar investigaciones cuantitativas en Ciencias Sociales.

otras entidades, siempre que utilicen el mismo sistema. En referencia a los registros, los datos pueden abarcar informaciones comerciales, de actividad profesional, etc.

- *Estadísticas de organizaciones*: en instituciones y organizaciones públicas y privadas es frecuente la existencia de registros y de estadísticas con fines documentales, registro u obligación. Los tipos de datos que podemos encontrar son tan diversos como sus orígenes, variando en cada institución a la que nos acerquemos en función de sus objetivos. Uno de los principales problemas que encontramos es llegar a saber de la existencia de estos datos, ya que no todas las organizaciones los hacen públicos. Si bien, si hablamos de información de las Administraciones Públicas, sí que es más común tener accesibilidad a estos datos. Es preciso conocer cómo se ha procedido a codificar variables, a construir índices, etc., dado que puede determinar la validez de la información que estamos obteniendo y cómo esta se adaptará a nuestra investigación.
- *Datos de expertos*: en algunos casos, los datos a los que queremos acceder no se encuentran disponibles en algún formato físico, y por ello es necesario acudir a ciertas personas que conocen dichos datos, útiles para nuestra investigación. Por ello, es esencial acceder a dichas personas a fin de obtener la información. A la hora de acceder a estas, la selección puede deberse a un doble motivo; por un lado, puede ser que dichas personas se encuentren implicadas en la acción, gestión o elemento que nos interesa conocer; o bien que dichas personas sean expertas conocedoras de los fenómenos que queremos analizar. Esta segunda opción es la más común. Si bien es cierto que, cuando se utiliza este tipo de datos, hemos de tener en cuenta que las informaciones pueden estar sesgadas por las experiencias vitales de los expertos. De igual manera, el hecho de no poder abarcar a todas las personas que podrían ser interesantes o de que la información que nos ofrezcan no sea del todo relevante para nuestra investigación es otro de los elementos a tener en cuenta.
- *Datos de encuestas*: nos detendremos poco en este punto, dado que será desarrollado posteriormente con gran detalle, dado que supone una de las principales fuentes de datos en Ciencias Sociales. Simplemente, hay que destacar la existencia de multitud de instituciones, empresas y organizaciones que realizan

encuestas de diferentes tipos (mercado, opinión...). En España, el más importante es el CIS,⁷ pero también existen otros regionales como el Ikuspegi⁸ (País Vasco). Además, tenemos otras internacionales, como la Encuesta Social Europea,⁹ la World Value Survey¹⁰ o el Eurobarómetro de la Unión Europea.¹¹

- *Datos documentales*: en esta categoría englobamos a todos aquellos documentos tanto de carácter público como privado que pueden ser empleados para obtener la información necesitada. No solo hablamos de informes, sino también de artículos de periódicos, epistolario privado, programas políticos, biografías, etc. Cualquier documento fidedigno que pueda ser utilizado para recopilar los datos necesarios para nuestra investigación es susceptible de ser empleado.

En cuanto a las técnicas de recogida, también encontramos diferencias, existiendo varias, entre las que destaca la encuesta, de la cual hablaremos a continuación. Aun así, merece la pena describir brevemente otro tipo de técnicas de recogida como las técnicas documentales, las entrevistas, los grupos de discusión y la observación. Así, comenzando por las primeras, son aquellas técnicas en las que se utiliza cualquier tipo de documento producido por una segunda persona u organismo y no por él mismo. La investigación bibliográfica estaría también aquí incluida. Sobre la base de esto, podemos asegurar entonces que toda in-

7. El CIS, cuyas siglas responden al Centro de Investigaciones Sociológicas, es el único centro público con datos de opinión pública de Europa y uno de los que mayor trayectoria tiene. Cuenta con investigaciones que van desde la opinión política hasta cuestiones sanitarias, de mercado o personales/familiares. Más información en su página web: <https://www.cis.es/cis/opencms/ES/index.html>

8. Ikuspegi – Observatorio Vasco de Inmigración inició su andadura en 2004 con vocación de construir una herramienta de utilidad pública para el conocimiento sistemático del fenómeno migratorio en el País Vasco. Cuenta con información estadística sobre diferentes cuestiones vinculadas a la inmigración. Más información en su página web: https://ikuspegi.eus/es/nuestras_publicaciones/barometros.php

9. La Encuesta Social Europea (ESS en sus siglas en inglés) es una encuesta internacional gestionada de forma académica que se realiza a lo largo y ancho de Europa desde 2001. Con carácter bianual, se realizan encuestas presenciales con muestras renovadas. Más información en su web: <http://www.europeansocialsurvey.org>

10. La información de este programa internacional de investigación puede comprobarse en este enlace: <https://www.worldvaluessurvey.org/wvs.jsp>

11. Todo lo relevante a esta institución puede consultarse aquí: <https://www.europarl.europa.eu/at-your-service/es/be-heard/eurobarometer>

investigación utiliza, en mayor o menor medida, la recogida de datos a través de técnicas documentales. Es importante, previo a la búsqueda de información a través de estas técnicas, pensar en diferentes palabras clave que puedan ser de ayuda a la hora de obtener la información o de realizar una búsqueda lo más efectiva posible (Vallés Martínez, 2000).

En cuanto a las entrevistas, la más utilizada, a pesar de existir diferentes técnicas, es la conocida como *entrevista en profundidad* o *semidirigida*. Esta técnica consiste en mantener una conversación larga (en torno a una hora y media o dos horas por sesión) con un individuo que posea información de relevancia para el investigador. Tras la entrevista o entrevistas, el investigador debe realizar un análisis de las respuestas obtenidas, ideas, elementos de dificultad... Para ello, aunque no es indispensable, se recomienda que se proceda a una grabación de audio y/o vídeo de cada entrevista. Con todo, muchas veces esto pueda suponer un problema en función de las individualidades a las que estemos entrevistando. Entre los elementos clave destaca la elaboración de una buena selección de sujetos a entrevistar, así como la preparación de un guion lo más completo posible (Denzin y Lincoln, 2005).

Sobre los grupos de discusión, cabe destacar que es una de las técnicas más valoradas entre los investigadores, junto con la que posteriormente destacaremos, la encuesta. Consiste en una reunión de un grupo de sujetos en una sala apta para este tipo de entrevistas (elementos técnicos para la grabación de audio y vídeo, mesa, sillas, posibilidad de reproducción de material audiovisual, cristal tintado desde el que poder realizar observación...), quienes, organizados por un moderador experto y conocedor del tema a tratar, debatirán sobre el tema propuesto y útil para la investigación en cuestión. Existen diferentes criterios a la hora de componer el grupo, si bien no hay unos más válidos que otros y dependerán del resultado que tratemos de obtener (Krueger, 1991).

Finalmente, respecto a la observación directa, son varias las técnicas que se engloba en este sistema, si bien las dos principales son las no reactiva y la participante. En la primera, el investigador se mantiene ajeno a los diferentes comportamientos y actitudes de los observados; por su parte, en la segunda, se persigue la interacción del investigador con los sujetos investigados, como forma de obtener la información más viable. La tempora-

lidad de este tipo de técnica de recogida de información es muy elevada, dado que es necesario que el investigador sea aceptado por el grupo (en el caso de la participante) y debe dilatarse en el tiempo para estar seguro de que estamos obteniendo la información deseada y que esta no se encuentra sesgada de ninguna manera. Resulta necesario llevar un seguimiento de las observaciones, una especie de cuaderno de campo o diario, a través del cual realizará una construcción de sus datos a partir de sus anotaciones y observaciones (Rodríguez, Gil y García, 1996).

Analizando la técnica de la encuesta y, en concreto, su instrumento capital representado como el cuestionario, cabe destacar que, como se ha indicado con anterioridad, esta técnica puede constituirse tanto en fuente primaria como secundaria (Fowler, 1995). Es una fuente primaria en la medida en que, para desarrollar nuestra investigación, en ocasiones las personas al frente diseñan y aplican un cuestionario en las diferentes formas en las que se puede implementar. Y, secundaria, en la medida en que existen diferentes organismos que ponen a disposición de la sociedad en general y de la comunidad científica en particular diferentes grupos de datos.

Esta técnica es ampliamente utilizada como procedimiento de investigación social, ya que permite obtener y elaborar datos de modo rápido y eficaz. Es, por excelencia, la estrategia cuantitativa asociada a los enfoques individuales que se centran en los sujetos (Rubio y Varas, 1997). En su acepción más simple, supone un método de recogida de información en el que esta se obtiene directamente de personas que han sido seleccionadas para establecer una base sobre la que puedan hacerse inferencias respecto a una población más amplia. Esta información se puede lograr a través de diferentes canales, aunque los principales suelen ser preguntando directamente a través de entrevistas personales o telefónicas, o bien haciendo que los sujetos rellenen cuestionarios enviados previamente. Gracias a esta facilidad en su implementación y, en consecuencia, a que la obtención de la información suele ser rápida y eficaz, la encuesta se ha caracterizado como una de las herramientas de investigación más utilizada en las ciencias sociales contemporáneas. Su formato y las múltiples posibilidades que permite en su puesta en práctica la han situado en la cúspide de las técnicas de recogida de datos (Pérez Castaños y Trujillo Cerezo, 2019).

Obtener una definición precisa de lo que es una encuesta puede resultar complejo, ya que se puede afirmar que existen tantas, prácticamente, como autores han pretendido definirla. Así, una de las más completas a nuestro juicio es la acuñada por Mayntz, Holm y Hübner (1975), que la caracterizaron como:

[la búsqueda] sistemática de información en la que el investigador pregunta a los investigados sobre los datos que desea obtener, y posteriormente reúne estos datos individuales para obtener durante la evaluación datos agregados. (p. 133)

Otros puntos de vista interesantes pueden traerse a colación a partir de las reflexiones de dos autores españoles. Si optamos por la definición de García-Ferrando (1993), entenderemos esta como una técnica cuantitativa que utiliza un conjunto de procedimientos estandarizados de investigación mediante los cuales se recogen y analizan una serie de datos de una muestra de casos representativa de una población o universo más amplio, del que se pretende explorar, describir, predecir y/o explicar una serie de características. Por su parte, para Sierra Bravo (1994), la observación por encuesta consiste igualmente en la obtención de datos de interés sociológico mediante la interrogación a los miembros de la sociedad, siendo el procedimiento sociológico de investigación más importante y el más empleado.

Más allá de los matices de los distintos autores al respecto, entre las características que prácticamente todos especifican de la investigación mediante encuesta, se podrían destacar las siguientes (Font Fábregas y Pasadas del Amo, 2016):

- La información se obtiene a partir de una observación indirecta de los hechos a través de las personas entrevistadas, en ese sentido, cabe la posibilidad de que la información obtenida no siempre refleje fidedignamente la realidad.
- La encuesta permite aplicaciones masivas que, mediante técnicas de muestreo adecuadas, pueden hacer extensivos los resultados a comunidades enteras.
- El interés del investigador no es el sujeto concreto que contesta el cuestionario, sino la representación de la población a la que en teoría pertenece; de ahí, como se ha mencionado, la necesidad de utilizar técnicas de muestreo apropiadas.

- Permite la obtención de datos sobre una gran variedad de temas, tanto en un plano objetivo, aludiendo directamente a las vivencias de los sujetos entrevistados, como subjetivo, referente a la forma en la que ve y percibe el mundo que le rodea.
- La información se recoge de modo estandarizado mediante un cuestionario (instrucciones iguales para todos los sujetos, idéntica formulación de las preguntas, etc.), lo que faculta hacer comparaciones entre todos ellos y, simultáneamente, su exportación y contraste con otras investigaciones similares.

Aparte de las características generales, hemos de tomar en consideración los términos principales que se utilizan en lo que a la investigación a través de encuesta se refiere. Así, en general, se llama *persona entrevistada* a quien responde a las preguntas de la encuesta; y por *cuestionario* se alude al instrumento elemental de toda encuesta que contiene el formulario definitivo o la batería de preguntas, especificaciones y claves a partir de las cuales se han de explicitar las respuestas (Sheatsley, 1983; Newell, 1993). Este sirve, en última instancia, de unidad elemental de recogida de información. Su finalidad es obtener de manera sistemática y ordenada toda la información sobre las variables que intervienen en una investigación.

Hay que destacar que existen diferentes tipos de encuesta a la hora de realizar una investigación cuantitativa con esta técnica; por ello, resulta capital conocer cuál se adapta mejor a cada uno de los escenarios en los que las diferentes investigaciones se pueden desempeñar. Por tipo de encuesta se alude a cuál será la vía de observación u obtención de los datos; concretamente, cómo se lleva a cabo el proceso administración del cuestionario. Una vez más, la selección más adecuada será la que mejor se adecue a las necesidades de información, teniendo en cuenta las siguientes características descritas por Díaz de Rada (2002):

- Nivel deseado de calidad de los datos.
- Presupuesto del que se disponga.
- Contenido del cuestionario (número y tipología de preguntas, alternativas de respuesta, etc.).
- Tiempo del que se dispone para su administración.
- Tipo del universo que se va a estudiar.

Por otra parte, hemos de destacar que los tres métodos de captar información mediante encuesta que se van a tratar aquí no son mutuamente excluyentes, sino que pueden combinarse y obtenerse así mejores resultados en el diseño de la investigación en función de las ventajas e inconvenientes que cada uno de ellos posee (Dillman, Smyth y Christian, 2009). Así, distinguimos entre encuestas *telefónicas*, las *autoadministradas* y las *personales*.

Las encuestas telefónicas tienen como objetivo fundamental conocer la opinión sobre algún tema, tendencia o preferencia concreta. Dado que el proceso está en la actualidad apoyado por diferentes programas de *software* informático, también son conocidas bajo el acrónimo CATI, cuyas siglas provienen de los términos ingleses *Computer Assisted Telephone Interview*.¹² Generalmente, la encuesta telefónica suele estar formada por cuestionarios cortos y preguntas directas, cuyas opciones de respuesta se encuentran limitadas (Lavrakas, 1993). Este tipo de encuestas han tenido una rápida implantación gracias a la generalización de teléfonos en domicilios particulares; sin embargo, desde la universalización de los dispositivos de comunicación portátiles, se ha complicado el acceso seguro a determinadas unidades de población (Pasadas, 2014). Entre las ventajas que encontramos al utilizar esta técnica destacamos la calidad de las respuestas obtenidas, ya que se puede no solo verificar la coherencia del entrevistado en tiempo real, sino también registrar posibles comentarios del entrevistador. Además, considerando las respuestas obtenidas, se pueden modificar las originalmente previstas en caso de existir problema de algún tipo. De la misma manera, la asistencia por ordenador permite añadir aleatoriedad a preguntas y respuestas de forma automática. Y, por último, otro de los elementos que optimiza la utilización de las encuestas telefónicas es la eliminación de costes de grabación de datos. Dado que el fichero se encuentra disponible casi inmediatamente, es posible comprobarlo de forma sencilla y, en consecuencia, se dota de mayor rapidez al desarrollo del trabajo de campo.

Por encuesta autoadministrada aquella en la que un entrevistado recibe un cuestionario y una carta de presentación o instrucciones a través de una vía indirecta de contacto (por ej.: co-

12. Literalmente, *entrevistas telefónicas asistidas por ordenador*.

recoordinario, electrónico o, cada vez más, mediante aplicaciones digitales) y responde a las preguntas cuando lo desea de forma autónoma, sin ayuda del investigador ni ningún otro intermediario. Únicamente debe seguir las instrucciones incluidas en el cuestionario y/o en la carta de presentación y posteriormente devolver el cuestionario cumplimentado al investigador (De Leeuw, 2004, p. 22). De esta manera, este tipo de herramienta prescinde, en términos generales, de la necesidad de personas que realicen las entrevistas, dado que el respondiente es quien completa el cuestionario o formulario de acuerdo con ciertas instrucciones, que deben ser sumamente claras y que se entregan a la par que el cuestionario. Dada la ausencia total o casi total del contacto personal en este tipo de encuestas, es imprescindible siempre que se tenga sumo cuidado con la manera en la que se confeccionan las preguntas y con la forma de presentar el cuestionario. Las preguntas deben ser lo más sencillas posibles, sin que esto afecte a la profundidad de lo que se quiere medir y, además, deben poder responderse con facilidad. Es necesario tomar en consideración que, al implementar este tipo de cuestionarios, puede que parte de la población quede fuera del rango del alcance de este. Este problema es más acuciante si cabe en el caso del cuestionario enviado a través de correo electrónico. En determinadas franjas de edad o contextos territoriales puede que no se disponga de las habilidades o de las capacidades técnicas necesarias para la cumplimentación de dichos cuestionarios por parte de los sujetos implicados (Alvira y Martínez Ramos, 1985). O, en otros casos, puede que obtengamos varias respuestas de las mismas personas. Este tipo de condicionantes deben ser tomados muy en cuenta a la hora de aplicar este tipo de estrategia.

La encuesta personal es quizás el método que goza de mayor popularidad y el que se ha utilizado con mayor profusión en la captación de información primaria, debido principalmente a las ventajas que presenta frente a los otros tipos de encuesta. Podemos definirlo como:

[...] el modo en el cual un entrevistador administra un cuestionario estructurado (o parcialmente estructurado) a un entrevistado en un periodo de tiempo determinado y en presencia (normalmente en el hogar) de este. (De Leeuw, 2004, p. 22)

En este tipo de encuesta se requiere de un equipo de entrevistadores convenientemente preparados, habiendo recibido formación específica referida a la naturaleza de la encuesta, la manera de administrar el cuestionario, algunos problemas con ciertas preguntas o cómo acceder a los sujetos entrevistados. El entrevistador debe acercarse al entrevistado y administrar la encuesta, por lo que el impacto que genera en los datos puede ser elevado. Su presencia, funciones, comunicación (tanto verbal como no verbal) y entonación pueden influir en las respuestas que da la persona que se somete al cuestionario. A través de este tipo de encuestas la calidad de la información recogida es alta, dado que se pueden incluir preguntas con respuesta abierta, en las que el entrevistado puede responder con la máxima libertad, así como el establecimiento de cuestionarios más largos que permitan cubrir diferentes puntos de vista del objeto de estudio (Díaz de Rada, 2001). De igual forma, la posibilidad de que existan evasivas o que el entrevistado quiera finalizar la entrevista antes de tiempo se reduce al enfrentarse al entrevistador cara a cara. Por último, este tipo de herramienta permite, con poca información, elaborar un marco muestral amplio y flexible, en el que se incluye información no solo del individuo entrevistado, sino también de la composición del hogar. Además, el error de este tipo de encuestas es preciso y conocido de antemano una vez se diseña. Cabe destacar, en concreto, que la tasa de cobertura de la encuesta, así como la tasa de respuesta que se obtiene, pueden ser catalogadas como de magnitud alta. Dicho en otros términos, los respondientes suelen realizar la totalidad del cuestionario y se puede llegar a prácticamente la totalidad de la muestra diseñada.

Para terminar este apartado, vamos a destacar cuáles son los puntos fuertes y débiles de este tipo de técnica (Díaz de Rada, 2005). Cuando nos aproximamos a la recogida de datos mediante encuesta, debemos tener en cuenta la existencia de una serie de ventajas e inconvenientes frente a otras estrategias de investigación social. Por un lado, en lo que a las ventajas se refiere, cabe destacar que es una importante y rica fuente de información si se realiza adecuadamente, dado que nos permite alcanzar una gran parte del universo y, además, obtener una información amplia que cubra la práctica totalidad del espectro del objeto de estudio (Cea D'Ancona, 1998; Díaz de Rada, 2001). Otro de los aspectos a tener en cuenta es su idoneidad para la indagación de

fenómenos cuantificables en torno a acciones, valores y/o preferencias (por ej.: la base social de apoyo a una determinada sanción o actitud policial), pues en los datos que se generan se puede conseguir una importante correspondencia con la realidad que se investiga o, lo que es lo mismo, una alta *validez científica*.¹³ Y, en esta línea, también nos permite estimar la precisión que tienen dichos datos respecto a lo que representan, lo que científicamente se conoce a grandes rasgos como *fiabilidad* (Fernández-Esquinas, 2003). Finalmente, entre las ventajas que plantea el análisis mediante encuesta, también se encuentra la posibilidad de realizar más fácilmente generalizaciones sobre poblaciones más amplias.

Por otro lado, en lo que a las desventajas se refiere, una de las más importantes que merece la pena destacar aquí es la relacionada con la capacidad de los respondientes a la encuesta (Díaz de Rada, 2002). En ocasiones, las personas que potencialmente deben contestar las preguntas del entrevistador puede que no posean capacidad suficiente para hacerlo. Esto puede deberse a una falta de conocimiento sobre los temas en cuestión, a falta de comprensión de las preguntas planteadas o de las respuestas ofrecidas, o incluso a barreras idiomáticas, entre otras cuestiones. En la práctica, esto imposibilita la aplicación óptima de la técnica y quizá nos deba llevar a plantearnos la necesidad de implementar una estrategia de otro tipo. Estos elementos o características, al igual que otros que mencionaremos más adelante en este apartado, se encuentran desarrollados con mayor profusión en diferentes epígrafes de este capítulo.

Otro de los grandes problemas a los que se enfrenta este tipo de investigación es su alto riesgo de subjetividad, es decir, que los valores del entrevistador o, incluso del investigador que diseña el instrumento de medición, se vean reflejados en la encuesta sesgando involuntariamente la recogida de información (a través de la formulación de las preguntas o en las respuestas enunciadas –Díaz de Rada, 2001–). Por otro lado, aspectos como la actitud, el volumen de voz, la indumentaria y la forma de presentarse de la persona entrevistadora pueden influenciar y deter-

13. El concepto de *validez científica* alude al grado en el que las medidas utilizadas se corresponden con los conceptos a reflejar (Manheim y Rich, 2002, p. 85). Será tratado en el último de los epígrafes de este capítulo.

minar las afirmaciones y respuestas. Finalmente, cabe destacar como desventaja o inconveniente la falta de aplicabilidad de este elemento ante el estudio de fenómenos que no se pueden cuantificar o que requieren de una complejidad terminológica en su respuesta. El siguiente cuadro ofrece un resumen más amplio de las ventajas e inconvenientes que han sido señaladas en este apartado, junto con otras destacadas por Cea D’Ancona (1998).

Tabla 1. Ventajas e inconvenientes de la encuesta

| Ventajas | Inconvenientes |
|--|---|
| <ul style="list-style-type: none"> • Permite abarcar un amplio abanico de cuestiones en un mismo estudio. • Facilita la comparación de resultados (al basarse en la estandarización y cuantificación de las respuestas). • Los resultados del estudio pueden generalizarse, dentro de los límites marcados por el diseño muestral efectuado. • Posibilita la obtención de una información significativa, cuando no acontezcan graves errores en su realización. • Ventaja económica: puede obtenerse un volumen importante de información a un mínimo coste (económico y temporal). | <ul style="list-style-type: none"> • No resulta adecuado para el estudio de poblaciones con dificultades para la comunicación verbal. • La información se restringe a la proporcionada por el individuo (a preguntas generalmente cerradas). • La presencia del entrevistador provoca efectos reactivos en las respuestas. • La carencia de referencias contextuales y vitales limita la interpretación de los datos de encuesta. • Acusa imprecisión para el estudio de la causalidad. • La existencia de obstáculos físicos (edificios vigilados, porteros automáticos, contestadores automáticos) dificultan el contacto con las unidades muestrales. • El desarrollo de una encuesta amplia resulta complejo y costoso (sobre todo en encuestas personales). |

Fuente: Cea D’Ancona (1998, p. 243).

2.5. Fundamentos estadísticos, análisis e interpretación de datos cuantitativos

A lo largo de este capítulo ya hemos expuesto cuáles son las características de las investigaciones cuantitativas, así como sus diferentes fases. Asimismo, sabemos cuáles son las diferentes técnicas de recogida y fuentes de datos existentes, así como hemos ampliado información en la técnica cuantitativa por excelencia, la encuesta. Ahora es el momento de, fruto de método estadístico que guía los análisis en las investigaciones de este tipo, explicar cómo

se realizan las diferentes operaciones estadísticas más habituales para el análisis, presentación y comentario de datos. Para ello, lo primero que debemos comentar es la inferencia estadística.

La estadística descriptiva y la teoría de la probabilidad van a ser los pilares de la inferencia estadística (García Esteban, 2011). La probabilidad y los modelos de distribución junto con las técnicas descriptivas constituyen la base de una nueva forma de interpretar la información suministrada por una parcela de la realidad que interesa investigar. En el siguiente esquema plasmado en la figura 2 se representa el tema a tratar y que será desarrollado a continuación.

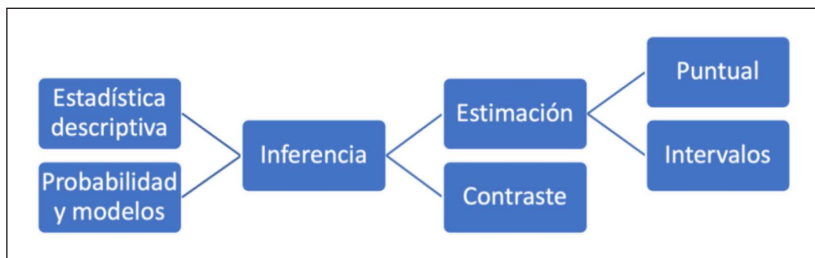


Figura 2. El proceso de inferencia estadística. Fuente: elaboración propia.

Los métodos básicos de la estadística inferencial son la estimación y el contraste de hipótesis, que juegan un papel fundamental en la investigación. En la mayoría de los análisis resulta imposible estudiar a todos y cada uno de los individuos de la población ya sea por el coste que supondría, o por la imposibilidad de acceder a ello. Mediante la técnica inferencial obtendremos conclusiones para una población no observada en su totalidad, a partir de estimaciones o resúmenes numéricos efectuados sobre la base informativa extraída de una muestra de dicha población. En definitiva, la idea es que, a partir de una población se extrae una muestra por algunos de los métodos existentes.¹⁴ De ella se obtienen datos numéricos que se van a utilizar para generar estadísticos con los que realizar estimaciones o contrastes poblacionales (Lynch, 2013).

Asimismo, como avanzábamos, la cantidad de elementos que integran la muestra (el tamaño de la muestra) depende de múlti-

14. Información más completa sobre las técnicas de muestreo en Rodríguez Osuna (2001, pp. 21 y ss.).

ples factores, como el dinero y el tiempo disponibles para el estudio, la importancia del tema analizado, la confiabilidad que se espera de los resultados, las características propias del fenómeno analizado, etc.

2.5.1. Análisis univariado, comentario y representación gráfica

Para poder realizar inferencia, hemos de conocer uno de sus pilares, la estadística descriptiva, también llamada *univariable*. Primero, debemos entender qué es: es aquella que se encarga de la exploración de una única variable (Carrasco Arroyo, 2012). Su objetivo es doble: por un lado, ordenar, elaborar y organizar la información recogida, y, por otro, resumir dicha información. Pero ¿cómo nos acercamos a esos objetivos? A través de la elaboración de distribuciones de frecuencias, de su representación gráfica y del cálculo de estadísticos univariados como las medidas de tendencia y variabilidad, como las que veremos en el este apartado.

Así, a la hora de comenzar a explorar una variable, lo primero que nos interesa saber es cuáles son sus elementos más importantes y la tipología de esta, dado que en función del tipo de variable que tengamos en nuestras manos (ordinal, numérica, nominal), podremos realizar una serie de estudios y no otros (Romo, 2014).

Lo primero que resulta capital comprobar es la distribución de las frecuencias de nuestra variable. Una distribución de frecuencias es una tabla que sustituye un conjunto de datos, difícilmente manejable por una presentación organizada de estos que facilite su interpretación (Lynch, 2013). Más en concreto, estamos hablando de una tabla con dos columnas, en donde a la izquierda se incluyen los valores o categorías de nuestra variable y a la derecha el número de veces que se repite cada una de ellas.

Encontramos cuatro elementos fundamentales:¹⁵

- *Frecuencia absoluta de un valor*: número de veces que se repite dicho valor.
- *Frecuencia relativa o porcentaje de un valor*: cociente entre la frecuencia absoluta y el tamaño del grupo (N total de casos) multiplicado por cien.

15. Las definiciones más completas se encuentran en Gil (2006).

- *Frecuencia absoluta acumulada de un valor*: número de veces que se repite en la distribución ese valor o cualquier otro valor inferior.
- *Porcentaje acumulado*: porcentaje de veces que se repite un valor o cualquiera de los valores inferiores.

Para poder obtener estos elementos, debemos acudir a algún programa estadístico ya sea más básico (Excel) o completos (SPSS, R, Stata). En este caso y, como ya se ha mencionado, debido a su amplio uso no solo en las disciplinas de las Ciencias Sociales, sino también en la docencia, se ha optado por el primero de los paquetes estadísticos más completos, el SPSS.

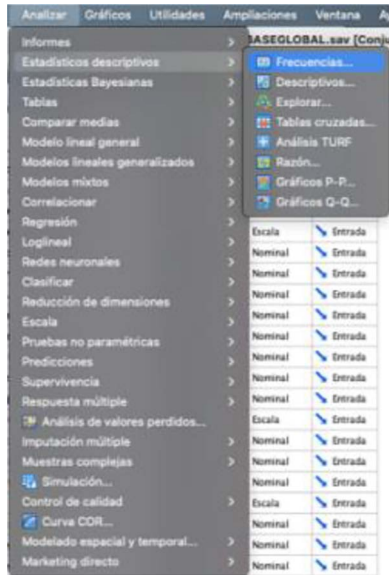


Figura 3. Proceso de solicitud de frecuencias con SPSS. Fuente: elaboración propia.

A través de SPSS, el procedimiento de solicitud es simple. Se debe acudir al menú *Analizar* y, de ahí, al desplegable de *Estadísticos descriptivos*. Se abrirá una nueva ventana emergente en donde a la izquierda se encontrarán todas nuestras variables, en un panel central en blanco será donde añadiremos las variables de las que queremos realizar nuestro análisis y a la derecha encontramos diferentes botones que nos ayudarán a solicitar exactamente los estadísticos univariados necesarios para el análisis.

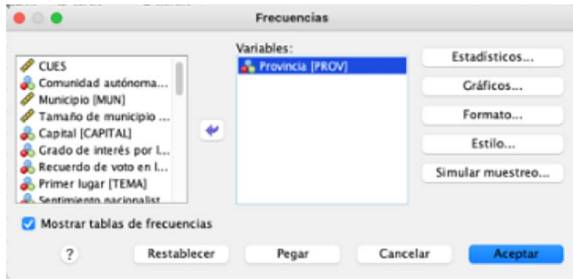


Figura 4. Ventana de frecuencias de SPSS. Fuente: elaboración propia.

En este caso, como simplemente nos interesa obtener la frecuencia, frecuencia relativa, frecuencia absoluta y porcentaje acumulado, no es necesario solicitar nada más en los botones de la derecha y, con seleccionar nuestras variables (de una en una, o en grupo a través de la tecla Ctrl del teclado) será suficiente. Para proceder a su inclusión en el panel central de la ventana iniciar se puede hacer con doble clic en la variable seleccionada o una vez seleccionada o seleccionadas las variables de interés, hacer clic en la flecha que se encuentra entre la columna de la izquierda y la central. Toda vez que ya disponemos de nuestras variables seleccionadas, únicamente debemos pulsar el botón Aceptar para poder generar nuestra o nuestras tablas de datos.

| Estadísticos | | | | | |
|---|---------------------------------------|------------|------------|-------------------|----------------------|
| Estatus socioeconómico de la persona entrevistada | | | | | |
| N | Válido | 21670 | | | |
| | Perdidos | 569 | | | |
| Estatus socioeconómico de la persona entrevistada | | | | | |
| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
| Válido | Clase alta y media alta | 4305 | 19,4 | 19,9 | 19,9 |
| | Clase media-media | 6359 | 28,6 | 29,3 | 49,2 |
| | Clase media-baja | 2695 | 12,1 | 12,4 | 61,6 |
| | Clase trabajadora/obrero/proletariado | 5255 | 23,6 | 24,3 | 85,9 |
| | Clase baja/pobre | 2712 | 12,2 | 12,5 | 98,4 |
| | Otras | 161 | ,7 | ,7 | 99,2 |
| | N.S. | 183 | ,8 | ,8 | 100,0 |
| | Total | 21670 | 97,4 | 100,0 | |
| Perdidos | N.C. | 569 | 2,6 | | |
| Total | | 22239 | 100,0 | | |

Figura 5. Resultados de frecuencias en SPSS. Fuente: elaboración propia.

Tras haber realizado los diferentes pasos, en la ventana de resultados del SPSS se nos arrojará una tabla como la que muestra la figura 5. Primero observamos en la parte superior izquierda una tabla de estadísticos que nos mostrará, para cada variable, cuál es el número de valores válidos y valores perdidos. Es decir, nos explicará cómo, en este caso para la variable estatus socioeconómico de la persona entrevistada, disponemos en nuestro análisis de 22.239 casos, de los cuales 21.670 son casos válidos y 569 perdidos.¹⁶

La segunda de las tablas que se muestran es, propiamente, la tabla de frecuencias. Esta se compone de cinco columnas, las cuales se dividen, de izquierda a derecha como siguen: categorías de la variable a analizar, frecuencia, porcentaje, porcentaje válido, porcentaje acumulado. La primera de las columnas especifica los valores o nombres que se ha asignado a cada una de las categorías de nuestra variable; en este ejemplo, cada nombre se corresponde con una horquilla de habitantes en las que cada municipio se engloba. La segunda de las columnas nos indica el número de respuestas que se hayan en cada una de estas horquillas de habitantes. Así, aquellas personas que se consideran de clase alta y media alta sumarían un total de 4305 respuestas, las personas que se consideran de clase media-media ascienden a 6359, y así sucesivamente. La tercera columna traslada estos datos a porcentajes sobre el total de respuestas obtenidas. De esta manera, el 4.305 de la primera columna se traslada a un 19,4%; es decir, esa cantidad de respuestas corresponden a un 19,4% del total. La cuarta columna nos muestra el porcentaje válido de nuestras respuestas que se encuadran en cada categoría de la variable objeto de estudio. Como hemos indicado, nuestra variable cuenta con 569 respuestas que no son válidas, por lo que los porcentajes se readecúan para adaptarse únicamente a las respuestas que hemos considerados como válidas. En nuestro ejemplo se aprecia cómo se han considerado como valores perdidos únicamente a aquellas personas que no responden a la pregunta (categoría *N.C.*), por lo que el 19,4% se convierte en un 19,9%, al eliminarse el 2,6% de respuestas no válidas. Finalmente, la quinta y última columna

16. Los valores más habituales en un análisis a través de encuesta suelen ser los «No Sabe» y «No Contesta»; es decir, opciones de respuesta que no aportan información de la variable o pregunta que se va a analizar. También recaería en esta categoría si en nuestra base de datos algunas de las casillas se encuentran sin información.

nos muestra el porcentaje acumulado; esto es, como, de forma agregada, se van sumando las diferentes respuestas para saber cuánto componen sobre el total. Así, el primer dato siempre coincidirá en la columna de porcentaje válido y en el porcentaje acumulado; sin embargo, si comparamos los valores de la segunda fila, vemos cómo en la cuarta columna se nos muestra un 29,3%, mientras que en la quinta el valor asciende a 49,2%. Este dato es fruto de la suma del 19,9% de la primera fila y del 29,3% de la segunda, indicando que 49,2% de las personas que respondieron a esta pregunta se sitúan entre las clases media-media y alta.

Sobre la representación gráfica, es necesario saber cuál es el mejor gráfico para seleccionar de cara a la presentación de los datos, como se verá posteriormente. Por ello, debemos seleccionar el más adecuado (Kastellec y Leoni, 2007). En función del tipo de exploración que tratemos de hacer, el tipo de gráfica que seleccionaremos variará. Se ofrece aquí una explicación resumida de cada uno de los gráficos más comúnmente usados a la hora de representar una variable.¹⁷

La primera de las representaciones gráficas más utilizadas es el Histograma. Consiste en una serie de rectángulos en donde el área de cada uno de ellos posee un tamaño proporcional a la frecuencia absoluta que representa. A través de esta representación se nos permite comprender una manera visual las desigualdades existentes entre las diferentes categorías existentes.

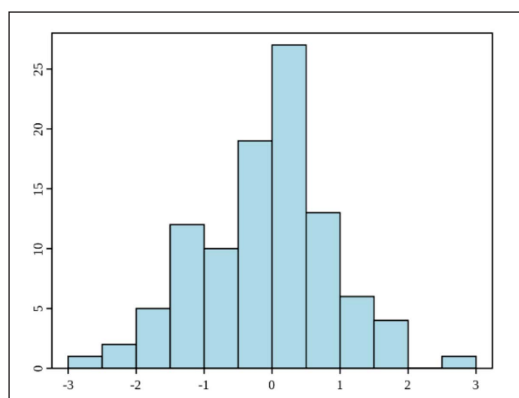


Figura 6. Histograma. Fuente: elaboración propia.

17. Información más completa y ejemplos visuales están disponibles en Alaminos (1993, p. 7 y ss.).

Por su parte, el polígono de frecuencias se utiliza uniendo los puntos de las partes superiores de los rectángulos del histograma, dibujando los puntos máximos de las frecuencias absolutas de los valores de la variable analizada. Este polígono puede ser suavizado para que se asemeje lo máximo posible a una curva y a la distribución normal, que posteriormente comentaremos. En tercer lugar, se hayan las ojivas, que representa las frecuencias acumuladas de la variable. Se puede representar tanto de mayor a menor como a la inversa. Así, el punto de cruce nos indicará la situación de la mediana, que explicaremos más adelante, y que divide la distribución en dos partes iguales.

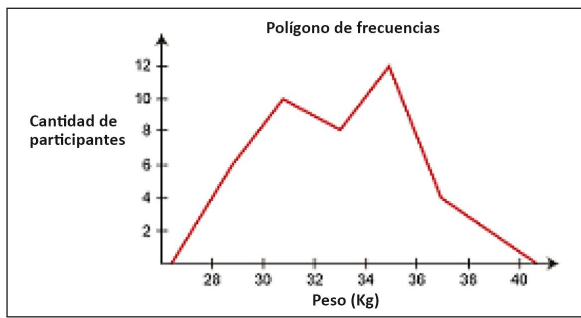


Figura 7. Polígono de frecuencias. Fuente: elaboración propia.

Los diagramas de barras o columnas simples nos permiten realizar comparaciones de un modo inmediato entre las diferentes categorías de una misma variable. Es muy sencillo de analizar, debido a la simplicidad de la comparación de alturas o longitudes de cada una de las barras.

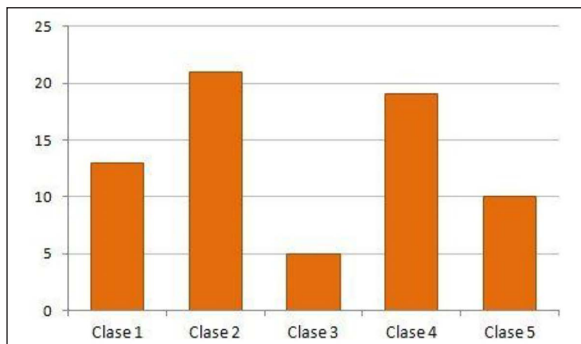


Figura 8. Diagrama de barras. Fuente: elaboración propia.

Finalmente, hablaremos de los gráficos de sectores, que se encuentran formados por un círculo dividido en varias partes en la que cada una de ellas representa una de las categorías de nuestra variable. Así, cada sector variará en tamaño en función de la amplitud de su frecuencia relativa o absoluta, en función del tipo que utilicemos.

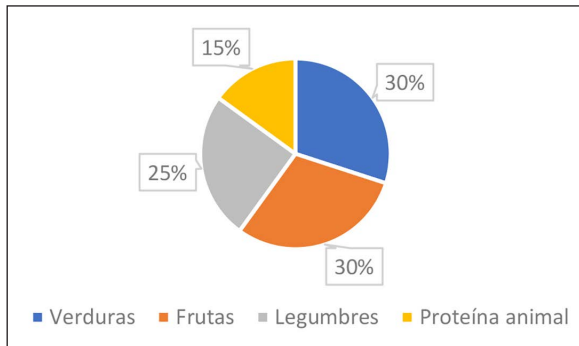


Figura 9. Gráfico de sectores. Fuente: elaboración propia.

Tras la comprobación de cómo representamos gráficamente los datos univariados, debemos entender cómo se distribuyen nuestros datos. Una de las finalidades de la estadística es la de describir un conjunto de observaciones en función de algunas magnitudes calculadas (estadísticos descriptivos), que expresan de forma resumida los rasgos más relevantes del grupo en observación.

Se hace, por tanto, conveniente una mayor reducción de los datos a una serie de «valores» que nos permitan por sí solos caracterizar dicho fenómeno. A estos valores los llamaremos *estadísticos* o *medidas descriptivas de los datos*.

De alguna forma, estos estadísticos nos resumen numéricamente algunas de las propiedades de las distribuciones de frecuencias: la tendencia central, la variabilidad o la forma. A estos estadísticos solemos denominarlos *medidas de tendencia central*, que se refiere a la magnitud general de las observaciones realizadas. Puede cuantificarse mediante unos índices numéricos conocidos como *índices de tendencia central* o *promedios*, porque pretenden ser síntesis de los valores de la variable (Minujin, D'Alessio y Oiberman, 1987).

Cada tipo de variables tiene su estadístico apropiado de tendencia central; así, para variables-atributos, se utilizará la moda,

mientras que para las variables cuantitativas se utilizarán tanto la mediana como la media aritmética (Manheim y Rich, 2002).

- *Moda*: es el valor de la distribución que más se repite. Para su cálculo solo es necesario saber contar.
- *Mediana*: es el valor central de la distribución. Puede definirse como *aquel valor de una distribución que deja por debajo de sí al 50% de los individuos que la forman*. Para su cálculo tendríamos que:
- *Media aritmética*: es el promedio de un conjunto de valores. Para su cálculo, se *suman todos los valores y se divide dicha suma por el número de ellos*. La media aritmética es el estadístico descriptivo de tendencia central más utilizado, entre otras razones por: la sencillez de su cálculo, la claridad de su interpretación y su precisión estadística, ya que utiliza toda la información obtenida.

Para poder saber cómo obtenemos estos valores, las figuras 10 y 11 nos detallan cómo se solicitan en SPSS. Los dos primeros pasos vienen descritos en las páginas anteriores y se pueden comprobar en las figuras 3 y 4. Así, una vez nos encontramos en el menú Estadísticos descriptivos de SPSS (figura 4), debemos apretar el botón Estadísticos, que se encuentra en la parte derecha, el cual nos abrirá un nuevo cuadro de diálogo que se muestra en la figura 10.

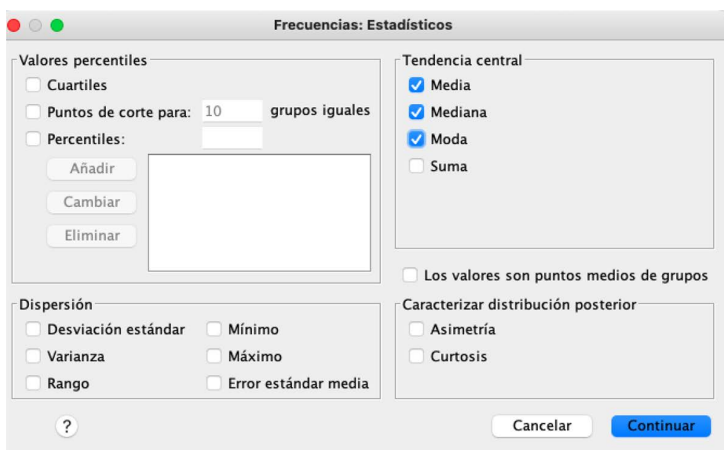


Figura 10. Opciones del botón Estadísticos Fuente: elaboración propia.

En este cuadro de diálogo nos aparecen multitud de opciones que podemos solicitar a SPSS para poder comprender y analizar un poco más profundamente la variable objeto de estudio, que, en este ejemplo, es la edad de la persona entrevistada. En la figura 9 se observa cómo se han marcado las opciones Media, Mediana, Moda de la parte superior izquierda del cuadro de diálogo. Este mismo cuadro de diálogo nos aportará la opción de seleccionar otros estadísticos descriptivos que comentaremos más adelante.

| Estadísticos | | |
|--|-----------------|--------------|
| Edad de la persona entrevistada | | |
| N | Válido | 22239 |
| | Perdidos | 0 |
| Media | | 50,77 |
| Mediana | | 50,00 |
| Moda | | 44 |

Figura 11. Resultados de las medidas de tendencia central. Fuente: elaboración propia.

Una vez que hemos marcado las medidas de tendencia central que nos interesan, volveríamos a la ventana mostrada en la figura 4 y clicáramos aceptar. En esta ocasión, aunque la tabla de resultados será similar a la vista en la figura 5, se le añaden a la que aparece en la parte superior de las informaciones los valores de media, mediana y moda. Tomando en consideración que la variable seleccionada es la edad de las personas entrevistadas, estos datos significan que: 1) habiendo sumado todos los valores y dividiendo entre el número total de casos, el valor promedio es de 50,77; 2) el valor que divide a nuestra muestra por la mitad es el 50; y 3) el valor más repetido en nuestra muestra es el 44. En términos de interpretación, decimos que la edad media de las personas entrevistadas es de 50,77 años, la mitad de la muestra se encuentra por debajo de los 50 años y la otra mitad por encima y, finalmente, que la edad más repetida en nuestra muestra es de 44 años.

Además de los aquí analizados, existen otros tipos de valores que sirven para reflejar la posición relativa que ocupa un individuo dentro de su grupo (Minujin, D'Alessio y Oiberman, 1987).

- *Percentiles* (centiles): son 99 puntuaciones de una distribución que dividen a esta en 100 partes iguales. El centil 23, cuyo símbolo es C23, será, por tanto, una puntuación que deja por debajo de sí al 23% de los individuos de una distribución, y, por consiguiente, dejará por encima al 77% restante.
- *Cuartiles*: son 3 puntuaciones de una distribución que dividen a esta en cuatro partes iguales. Su símbolo estadístico es Q1, para el cuartil 1; Q2, para el segundo cuartil y Q3 para el tercero.
- *Deciles*: son 9 puntuaciones de una distribución que dividen a esta en 10 partes iguales. Su símbolo estadístico es D1, para el decil 1, y así sucesivamente.

Para poder calcular estos estadísticos nuevos en SPSS debemos seguir los mismos pasos que hemos visto en las figuras 3, 4 y 10. De nuevo en el botón Estadísticos, que aparece cuando queremos solicitar una frecuencia, se nos abrirá el cuadro de diálogo que aparece en la figura 12 y deberemos seleccionar aquellos elementos que nos interesan más.

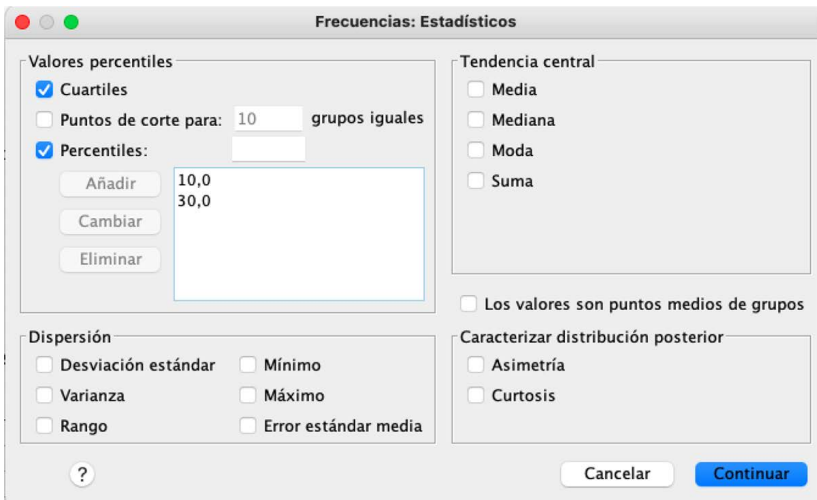


Figura 12. Cuartiles y percentiles en el botón Estadísticos. Fuente: elaboración propia.

Si en la figura 10 únicamente seleccionábamos las medidas de tendencia central más habitual, en este caso nos hemos mo-

vido hacia la parte izquierda del cuadro, marcando que queremos la opción de los cuartiles y la de percentiles. En este ejemplo, de nuevo con la edad de las personas entrevistadas, hemos indicado que queremos los percentiles 10 y 30 únicamente. Para poder indicar tanto los percentiles como los cuartiles que nos interesan, una vez que hemos marcado la opción Percentiles, debemos escribir el número de percentil en el cuadro justo a la derecha de la palabra y clicar en la opción Añadir. Si hemos cometido algún error al introducir el número, procederemos a editar seleccionando el número del percentil erróneo y marcaremos la opción Cambiar, lo cual nos permitirá editarlo. Si, por el contrario, queremos incluir alguno de los percentiles que hemos indicado, seleccionaríamos ese percentil y marcaríamos la opción Eliminar.

| Estadísticos | | |
|--|-----------------|--------------|
| Edad de la persona entrevistada | | |
| N | Válido | 22239 |
| | Perdidos | 0 |
| Percentiles | 10 | 26,00 |
| | 25 | 37,00 |
| | 30 | 40,00 |
| | 50 | 50,00 |
| | 75 | 64,00 |

Figura 13. Resultados de los cuartiles y percentiles. Fuente: elaboración propia.

Así, una vez que hemos seleccionado todos los elementos que son de interés, le diríamos a SPSS que ejecute las frecuencias y obtendríamos una primera tabla similar a la mostrada en las figuras 5 y 11, pero con más información. Debajo de los valores perdidos quedarían registrados los percentiles y cuartiles que le hemos solicitado. Así, comenzando con los percentiles solicitados, SPSS nos indica que el percentil 10 se encuentra en el valor 26 y el 30 en el valor 40. O, lo que es lo mismo, que el 10% de los casos se encuentran antes de los 26 años y el 30% hasta los 40 años. En la medida en que los diferentes cuartiles se corresponden también con los percentiles 25, 50 y 75, SPSS lo refleja

de esa manera y no a través del sistema de Q1, Q2 o Q3. Por ello, según los datos de la figura 12, vemos cómo el 25 % de los casos (el primer cuartil) se encuentra hasta los 37 años de edad; el 50 % (el cuartil dos) se haya en los 50 años (recordemos que era el valor de nuestra mediana) y el tercer cuartil o el 75 % de los casos los marcan los 64 años de edad.

Junto con la distribución, encontramos medidas complementarias que nos ayudan a comprender cómo se estructuran los datos dentro de cada variable. Estas reciben el nombre de *medidas de variabilidad* y miden el grado medio de alejamiento o separación de las observaciones respecto a las medidas de posición central.

- *Varianza*: es la más utilizada, fundamentalmente por su facilidad de cálculo, su buena prestación a los desarrollos algebraicos y sus propiedades estadísticas. La varianza valdrá cero si todos los valores de la variable son iguales a la media aritmética, y será mayor cuanto más alejados estén los valores de ella. Cuanto mayor sea la varianza de una distribución, los sujetos serán más heterogéneos respecto a ella, más dispersas están las observaciones unas de otras (Gil, 2006).
- *Desviación típica*: es la raíz cuadrada de la varianza. *Varianza* y *desviación típica* son términos relativos. Su interpretación no se puede hacer de forma absoluta, sino haciendo referencia a la variabilidad de la misma variable en varios grupos o a la variabilidad de distintas variables (Anduiza, Crespo y Méndez, 2011).

Así, el último de los análisis de medidas de tendencia central de un análisis univariado queda reflejado en las figuras 14 y 15. De nuevo, partimos de los pasos que hemos visto en las figuras 3, 4, 10 y 12 para poder abrir el cuadro de diálogo de los estadísticos y proceder a la obtención de la desviación típica y la varianza, tal y como se muestra en la figura 14.



Figura 14. Desviación típica y varianza en SPSS. Fuente: elaboración propia.

Esta vez, para poder solicitar los estadísticos relativos a la desviación típica y a la varianza, debemos fijarnos en la parte inferior izquierda del cuadro de diálogo. Ahí, las dos primeras opciones son las que nos interesan. Existen otros valores interesantes como pueden ser el mínimo (valor mínimo de nuestra variable) el máximo y el error estándar entre otros, para medirlos, seleccionaremos los dos primeros y volveremos a ejecutar las frecuencias para obtener los valores.

| Estadísticos | | |
|---------------------------------|----------|---------|
| Edad de la persona entrevistada | | |
| N | Válido | 22239 |
| | Perdidos | 0 |
| Desv. Desviación | | 17,714 |
| Varianza | | 313,796 |

Figura 15. Resultados de la desviación típica y la varianza. Fuente: elaboración propia.

Tal y como se aprecia en la figura 15, de nuevo la tabla de estadísticos cambia, manteniendo las dos primeras filas idénticas a lo analizado en otros momentos, y añadiendo en este caso los

valores de la desviación típica (denominada *Desv. Desviación*) y la varianza. El primero de los valores es de 17,714 indicando que la variable tiene una variabilidad en ese sentido. Igualmente, la varianza nos arroja un valor de 313,796, el cual, al ser distinto de 0, nos informa de que efectivamente nuestros datos no son todos idénticos y que tienen una variabilidad elevada. Obviamente, al tratarse de una variable como la edad, los valores posibles oscilan entre 18 y 99 años, por lo que nos indica la heterogeneidad de las respuestas obtenidas.

Finalmente, estas puntuaciones las utilizamos para generar valores estandarizados, también conocidos como *típicos* (Kaltenbach, 2012). Estadísticamente hablando, la puntuación típica de un determinado valor indica el número de desviaciones típicas que dicho valor se separa de la media del grupo de valores. Las puntuaciones típicas son el resultado de dividir la diferencia (distancia) de cada valor respecto a su media aritmética entre la desviación típica. Este proceso se llama *tipificación*.

Las puntuaciones típicas se representan por z :

$$Z = \frac{X_i - \bar{X}}{\sigma}$$

Estas nuevas puntuaciones cumplen los tres requisitos necesarios para comparar puntuaciones de distintos grupos o variables. Las puntuaciones típicas tienen la misma media ($= 0$), son puntuaciones con la misma dispersión o variabilidad ($= 1$) y tienen la misma unidad de medida (el valor de la desviación típica).

Estas puntuaciones típicas guardan una estrecha relación con un tipo de modelo de distribución llamado *curva normal*, *distribución normal* o, familiarmente, *campana de Gauss* (Wonnacott y Wonnacott, 2004). Este tipo de distribución dice que los datos, en términos generales, se distribuyen de una manera de campana, englobándose un mayor número de datos en los valores centrales y en torno a la media, para ir decayendo conforme nos acercamos a los extremos.

2.5.2. El análisis bivariado y su presentación gráfica

Es aquella parte de la estadística descriptiva que se ocupa del análisis estadístico de dos variables. Su objetivo es estudiar la re-

lación estadística entre variables (Manheim y Rich, 2002). Dicho objetivo, dependiendo del tipo de variables involucradas, se puede cubrir a través de diferentes técnicas y herramientas. A la inversa de lo que hicimos en el apartado anterior, comenzaremos hablando de la representación gráfica de dos variables para, posteriormente, desmenuzar las posibilidades del SPSS para obtener la información de nuestro interés.

En cuanto a las representaciones gráficas, a continuación, describimos algunos de los modelos más útiles e importantes a la hora de representar gráficamente exploraciones de dos o más variables.¹⁸ El primero es el diagrama de barras o columnas múltiples. Este tipo de gráfica permite, junto con el siguiente, conocer la distribución de dos o más variables o de su entre sus diferentes categorías. El diagrama de barras o columnas compuestas, el cual se puede disponer de dos formas, aquellos que poseen base 100% y los que no hacen. Comenzando por estos últimos, la representación se realiza de las frecuencias de cada una de las categorías de nuestras variables que queremos comparar, acumulándolas. De esta manera, nos permite conocer visualmente las diferencias de nuestras categorías en función de su número de casos; la limitación que poseen es que es necesario que no se den demasiadas separaciones, dado que, de lo contrario, resultaría casi imposible su lectura. Por otra parte, los diagramas compuestos que operan en base 100% se diferencian de los anteriores en la medida en que operan a partir de las frecuencias relativas. Así, podemos comparar el peso de cada categoría en las diferentes variables que procederemos a analizar o comparar.

18. Al igual que hicimos en el apartado de representación gráfica de una variable, recomendamos la lectura de Alaminos (1993, pp. 37 y ss.) para explicaciones más completas y ejemplos visuales que, por motivos de espacio, no hemos podido incluir aquí.

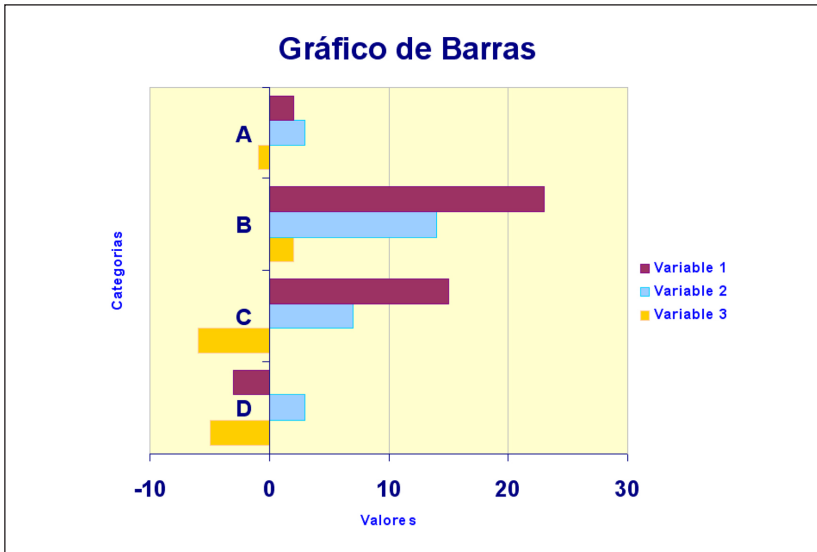


Figura 16. Diagrama de columnas simples. Fuente: elaboración propia.

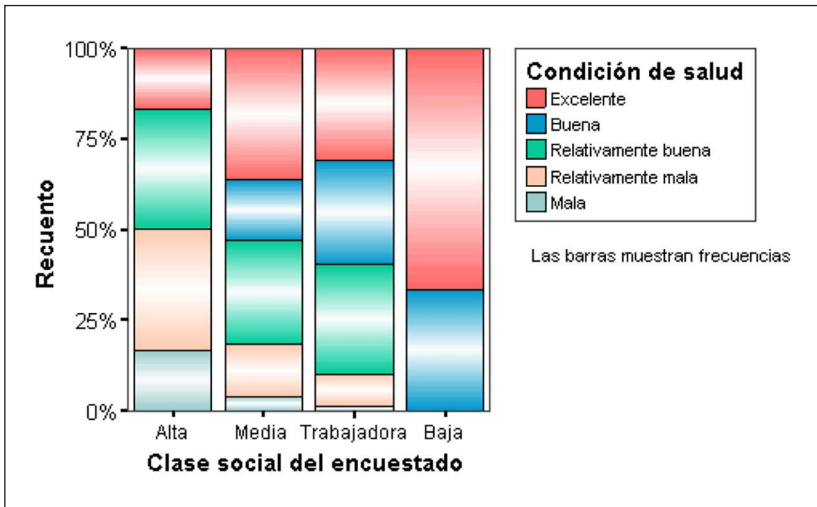


Figura 17. Diagrama de columnas compuestas. Fuente: elaboración propia.

El tercero de los gráficos es el de líneas simples. Estas representaciones son útiles para recoger la evolución de una unidad a lo largo del tiempo y especialmente óptimas para el análisis de series temporales de datos. Sin embargo, es necesario utilizar de forma correcta estas líneas, dado que, por ejemplo, existen ín-

dices que, a pesar de que puedan ser explicativos, la utilización de sus números inversos pueda ser más visual (por ej.: tasa de anal-fabetos o de alfabetización). Una variación de este es el de líneas múltiples, pudiendo representar la evolución temporal de varias variables, lo cual nos puede permitir comprender de forma visual si existe alguna relación entre las variables (como que todas aumenten a la par, o algunas mantengan una tendencia similar). A la hora de componer este tipo de gráficos, debemos tener en cuenta la capacidad visual de estos, y no incluir demasiadas variables en el análisis, dado que, de lo contrario, nos encontramos con un espacio repleto de líneas, que obstaculizará su lectura.

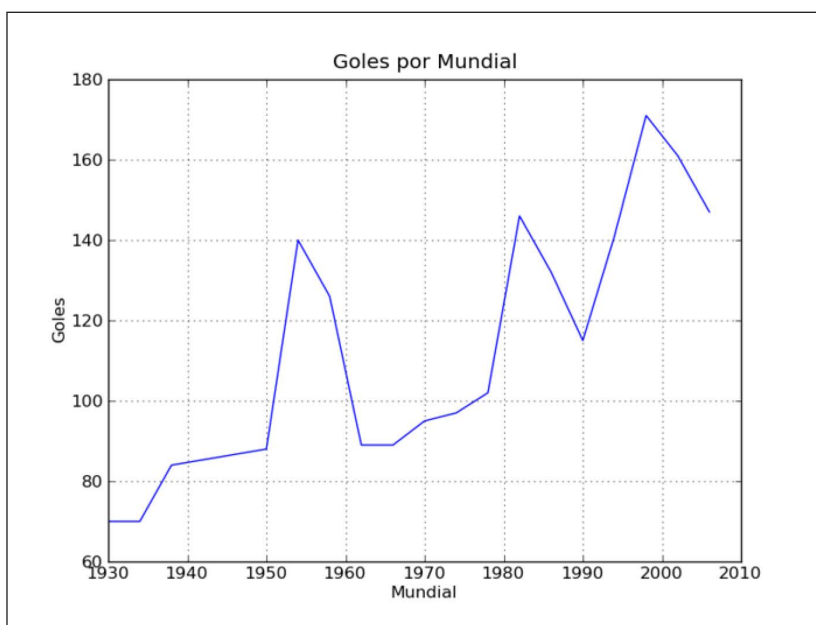


Figura 18. Gráfico de líneas simples. Fuente: elaboración propia.

Para acabar, hablaremos de las líneas compuestas o estratos, que resultan muy útiles cuando estamos tratando de comparar datos que, a su vez, forman parte de un agregado (por ej.: presupuestos generales del estado). Esto nos permite representar el peso específico de cada uno de ellos, además de su evolución temporal. De esta manera, cada estrato se encuentra superpuesto al anterior, siendo el total igual a la suma de los estratos para cada uno de los elementos analizados o periodos temporales.

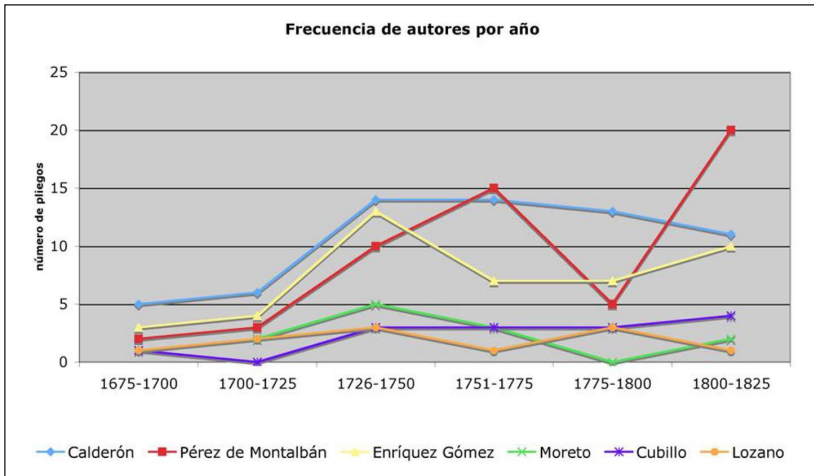


Figura 19. Gráfico de líneas compuestas. Fuente: elaboración propia.

Aparte del formato de gráfico, cuando queremos presentar dos o más variables encontramos otras soluciones igual de visuales y con capacidad analítica. Así, se denomina *tabla de contingencia* a la tabla que representa una distribución bidimensional. En la primera fila y columna se escriben las definiciones de cada una de las modalidades asociadas a los atributos. La exploración de datos a través de las tablas de contingencia es una de las fórmulas más comunes de realizar análisis de dos o más variables en ciencias sociales. De esta manera, se puede conocer la relación entre dos variables nominales u ordinales. Según el modelo más utilizado, en las filas, a la izquierda, se incluirá a nuestra variable dependiente, mientras que, en las columnas, arriba, se explicitarán las categorías de nuestra variable independiente (Tomeo y Uña, 2009).

Los resultados nos indican qué porcentaje de cada categoría se encuentra representado en el otro. Por ello, debemos tener cuidado ante cómo presentamos nuestros porcentajes, si por filas o por columnas. Así, los primeros nos indicarían cuántos casos dentro de cada categoría de nuestra variable dependiente se engloban en cada categoría de la independiente. Por el contrario, si decidimos mostrar los porcentajes por columnas, el resultado que arrojaría se correspondería con el inverso; es decir, con cuántos casos de cada categoría de nuestra variable independiente se engloban en cada una de las categorías de nuestra variable dependiente.

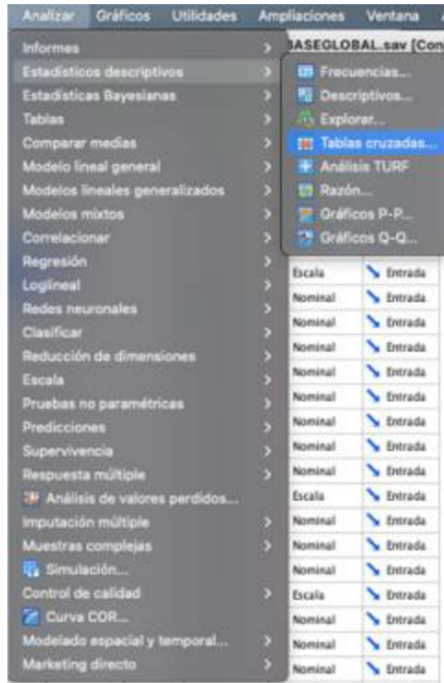


Figura 20. Proceso de obtención de tablas cruzadas. Fuente: elaboración propia.

Para poder obtener un cruce bivariado en SPSS, hemos de acudir al menú *Analizar*, allí seleccionar *Estadísticos descriptivos* y, del menú desplegable que se abre, buscar la cuarta opción, *Tablas cruzadas*, tal como se muestra en la figura 20. Una vez que lo hayamos seleccionado, se nos abrirá un menú como el que aparece en la figura 21, en donde encontraremos una columna con todas nuestras variables a la izquierda, una columna central con tres cajas, y una columna a la derecha con diferentes botones que obtendrán diferentes elementos extra que iremos comentando más adelante. Por ahora, en las siguientes figuras trabajaremos con nuestras variables, los cuadros centrales y el tercero de los botones que aparece en la columna derecha y que se denomina *Casillas*. Con estos elementos seremos capaces de obtener los primeros resultados preliminares bivariados. En este ejemplo trabajaremos con las variables tamaño de hábitat (tamaño del municipio en el que residen las personas que responden a la encuesta) y religión de las personas entrevistadas (bajo el nombre de *Católicos*, simplificadas en *Católicos* y *No católicos*).

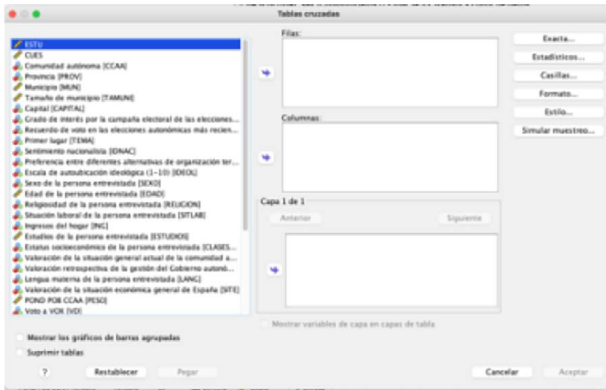


Figura 21. Menú inicial de Tablas cruzadas. Fuente: elaboración propia.

Para poder realizar una tabla, tenemos que seleccionar, de nuestra lista de variables que aparece en la columna de la izquierda de la figura 21, aquellas variables que buscamos analizar. Las convenciones académicas estableces que, a la hora de presentar las variables en una tabla, la variable dependiente suele presentarse en las filas y la independiente en las columnas. Cabe destacar que SPSS generará una tabla diferente para cada par de variables que se incluyan en el cuadro de Filas y en el de Columnas del bloque central del cuadro de diálogo. De este modo, si ponemos una variable dependiente en filas y tres variables independientes en columnas, nos generará tres tablas. Pero si ponemos dos variables dependientes en filas, el número de tablas será de seis.

| Resumen de procesamiento de casos | | | | | | |
|-----------------------------------|--------|------------|---------------|------------|-----------|------------|
| | Válido | | Casos Perdido | | Total | |
| | N | Porcentaje | N | Porcentaje | N | Porcentaje |
| CATOLICOS * Tamaño de municipio | 22238 | 100,0% | 1,075 | 0,0% | 22239,075 | 100,0% |

| Tabla cruzada CATOLICOS*Tamaño de municipio | | | | | | | | | |
|---|--------------|----------------------------------|---------------------------|----------------------------|-----------------------------|------------------------------|--------------------------------|-------|-----------------------------|
| Recuento | | Tamaño de municipio | | | | | | Total | |
| | | Menos o igual a 2.000 habitantes | 2.001 a 10.000 habitantes | 10.001 a 50.000 habitantes | 50.001 a 100.000 habitantes | 100.001 a 400.000 habitantes | 400.001 a 1.000.000 habitantes | | Más de 1.000.000 habitantes |
| CATOLICOS | No católicos | 820 | 2002 | 3547 | 1578 | 3141 | 1782 | 2616 | 15486 |
| | Católicos | 602 | 1053 | 1619 | 778 | 1347 | 810 | 543 | 6752 |
| Total | | 1422 | 3055 | 5166 | 2356 | 4488 | 2592 | 3159 | 22238 |

Figura 22. Resultados generales de tablas cruzadas. Fuente: elaboración propia.

Una vez que hemos incluido en las filas la variable Católicos y en las columnas el tamaño de hábitat, SPSS nos arroja los siguientes resultados. Al igual que sucedía en los procesos univariados, lo primero que se nos ofrece es una tabla de resumen, que nos indica que se han utilizado 22.238 casos válidos y que 1 ha quedado fuera. Con ello, podemos ver, en términos absolutos, el número de personas en nuestra muestra que se consideran No católicos y Católicos, y cómo estos se distribuyen por los diferentes municipios. Así, para el caso de los no católicos residentes en municipios de menos de 2000 habitantes comprobamos que la muestra es de 820 casos; si queremos comprobar los católicos que residen en grandes ciudades de más de 1.000.000 de habitantes, debemos acudir a la segunda fila y a la última columna, para comprobar como el número asciende a 543.

Estos resultados, como decíamos, son valores absolutos y no permiten realizar una comparación adecuada en términos de, por ejemplo, saber si existe un porcentaje mayor de católicos en alguno de los hábitats o si, por el contrario, no existen diferencias entre los grupos religiosos dependiendo del tamaño del municipio de residencia. Para poder hacerlo, hemos de acudir al botón Casillas, el tercero de los que aparecen en la sección derecha de la figura 21.

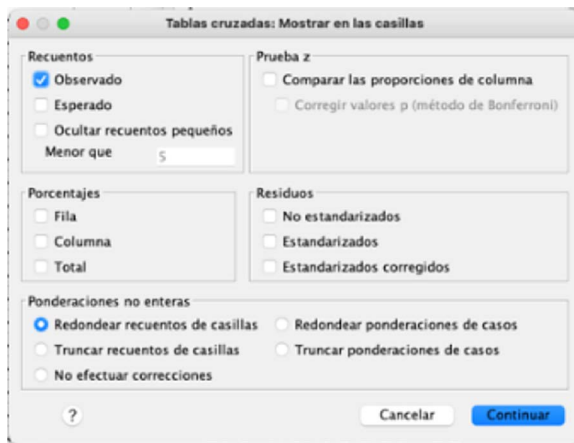


Figura 23. Menú Casillas de Tablas cruzadas. Fuente: elaboración propia.

Al hacer clic en el botón Casillas, se abrirá el menú que aparece descrito en la figura 23. A través de este se puede obtener una

cantidad de información abrumadora de las variables objeto de análisis en este ejemplo; sin embargo, en términos prácticos nos vamos a centrar en las opciones que aparecen en la sección intermedia de la columna de la izquierda. Esta sección se rotula como «Porcentajes» y disponemos de tres opciones para marcar que son Fila, Columna y Total. Estas tres opciones pueden marcarse de forma individual o varias a la vez, en función de los intereses de las personas que se encuentren realizando la investigación. En las figuras 24 y 25 se han seleccionado de forma individualizada la opción Filas y la opción Columnas, respectivamente.

| | | Tabla cruzada CATOLICOS*Tamaño de municipio | | | | | | | | |
|-----------|--------------|---|---------------------------|----------------------------|-----------------------------|------------------------------|--------------------------------|-----------------------------|-------|--------|
| | | Tamaño de municipio | | | | | | | | |
| | | Menos o igual a 2.000 habitantes | 2.001 a 10.000 habitantes | 10.001 a 50.000 habitantes | 50.001 a 100.000 habitantes | 100.001 a 400.000 habitantes | 400.001 a 1.000.000 habitantes | Más de 1.000.000 habitantes | Total | |
| CATOLICOS | No católicos | Recuento | 820 | 2002 | 3547 | 1578 | 3141 | 1782 | 2616 | 15486 |
| | | % dentro de CATOLICOS | 5,3% | 12,9% | 22,9% | 10,2% | 20,3% | 11,5% | 16,9% | 100,0% |
| | Católicos | Recuento | 602 | 1053 | 1619 | 778 | 1347 | 810 | 543 | 6752 |
| | | % dentro de CATOLICOS | 8,9% | 15,6% | 24,0% | 11,5% | 19,9% | 12,0% | 8,0% | 100,0% |
| Total | | Recuento | 1422 | 3055 | 5166 | 2356 | 4488 | 2592 | 3159 | 22238 |
| | | % dentro de CATOLICOS | 6,4% | 13,7% | 23,2% | 10,6% | 20,2% | 11,7% | 14,2% | 100,0% |

Figura 24. Resultado con porcentajes por filas. Fuente: elaboración propia.

Al haber seleccionado la opción Filas, debemos leer nuestros datos de izquierda a derecha, sabiendo que siempre se analizarán los resultados a partir de la categoría de nuestra variable dependiente en este caso. Por ello, los resultados que aquí se ofrecen toman en consideración el total de los católicos, por un lado, y no católicos, por otro, y buscan comprobar como cada uno de ellos se distribuye en las diferentes secciones de hábitat existentes.

Así, para poder interpretar los resultados de la figura 24, debemos decir que el 5,3% de las personas que no se consideran católicas residen en municipios menores de 2000 habitantes, el 22,9% lo hace en municipios de entre 10.000 y 50.000 habitantes y un 11,5% los hacen en municipios de entre 400.000 y un millón de habitantes. Estos datos ya pueden ser comparables con los datos de católicos, pudiendo comprobar de manera indiciaria si, por ejemplo, el tamaño del hábitat influye en la religiosidad. En el caso de los municipios de entre 2000 y 10.000 habitantes, el porcentaje de no católicos es de 12,9%, mientras que el porcentaje de católicos es de 15,6%; entre el siguiente tramo de habitantes (entre 10.000 y 50.000), el porcentaje de católicos es de 1,1% puntos mayores (22,9% frente a 24%) y en las urbes

de más de un millón de habitantes los católicos superan a los no católicos por apenas un 0,5 %. Con estos resultados, sí que podemos proceder a, como indicábamos, una comparativa entre categorías.

| | | | Tamaño de municipio | | | | | | | |
|-----------|--------------|---------------------------------|----------------------------------|---------------------------|----------------------------|-----------------------------|------------------------------|--------------------------------|-----------------------------|--------|
| | | | Menos o igual a 2.000 habitantes | 2.001 a 10.000 habitantes | 10.001 a 50.000 habitantes | 50.001 a 100.000 habitantes | 100.001 a 400.000 habitantes | 400.001 a 1.000.000 habitantes | Más de 1.000.000 habitantes | Total |
| CATOLICOS | No católicos | Recuento | 820 | 2002 | 3547 | 1578 | 3141 | 1782 | 2616 | 15486 |
| | | % dentro de Tamaño de municipio | 57,7% | 65,5% | 68,7% | 67,0% | 70,0% | 68,8% | 82,8% | 69,6% |
| CATOLICOS | Católicos | Recuento | 602 | 1053 | 1619 | 778 | 1347 | 810 | 543 | 6752 |
| | | % dentro de Tamaño de municipio | 42,3% | 34,5% | 31,3% | 33,0% | 30,0% | 31,3% | 17,2% | 30,4% |
| Total | | Recuento | 1422 | 3055 | 5166 | 2356 | 4488 | 2592 | 3159 | 22238 |
| | | % dentro de Tamaño de municipio | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |

Figura 25. Resultados con porcentajes por columnas. Resultado con porcentajes por filas. Fuente: elaboración propia.

Si, por el contrario, la comparativa la queremos hacer por diferentes tamaños de hábitat, en la figura 25 se arrojan los resultados a partir de estas categorías. Así se han tomado como referencia el total de personas que se encuentran en cada sección poblacional y se calcula el peso relativo de cada una de las categorías de nuestra variable dependiente en ellas. De esta manera, del total de respondientes que residen en municipios de menos de dos mil habitantes, un 57,7 % se consideran no católicos, mientras que un 42,3 % se consideran católicos. Si comprobamos a los residentes en ciudades de más de un millón de habitantes, el porcentaje de no católicos alcanza el 82,8 %, mientras que el de personas que sí que se definen como católicos se queda en un 17,2 %.

Es importante no calcular este tipo de tablas con variables con muchas categorías o, al menos, no presentarlas así en un posterior informe, dado que la función principal es la de comprender de manera visual si existe algún tipo de relación entre las variables incluidas. Además, también resulta relevante incluir el número de casos total y en cada una de las celdas.

Cuando el interés se dirige a conocer si un fenómeno está definido simultáneamente por distintas variables, la estadística pone a nuestra disposición distintas técnicas para establecer si existe o no correlación entre las variables. Así, según Cortés y Rubalcava (1987), la asociación más simple y frecuente entre dos variables es por contraposición a la independencia estadística. Para esto se

desarrollan coeficientes que tratan de resumir la distribución conjunta a través de magnitudes rápidamente interpretables.

En el análisis de correlación entre dos variables es necesario distinguir aspectos distintos que pueden ser objetivo de una hipótesis (Lynch, 2013):

- La existencia de una asociación simplemente verificada a través del rechazo de la hipótesis nula que afirma independencia estadística.
- La magnitud de la asociación existente entre dos variables a partir de una escala fácilmente interpretable (muy fuerte/ fuerte/moderada/ débil/ despreciable). Esta noción de magnitud de asociación supone que antes se ha descartado la hipótesis nula de la independencia estadística.
- El sentido de la asociación entre las variables. Por ejemplo, afirmando que cuando una se incrementa otra disminuye.

Lógicamente, el esquema expone que las medidas que permiten analizar la forma de la relación incorporan todas las propiedades de las medidas que permiten analizar el sentido, la magnitud y la existencia (Gil, 2006).

Desde el punto de vista estadístico, se suele sostener que los coeficientes más apropiados por versátiles y elocuentes, tienen las siguientes propiedades:

- No están afectados por el N total de la distribución. Es decir, están normalizados.
- El valor 0 indica la inexistencia de una relación o más estrictamente, independencia estadística.
- Existe un valor máximo teóricamente establecido, el mismo para cualquier distribución, que indica una relación de asociación o correlación perfecta.
- En el caso de que el objetivo sea un análisis del sentido, también existirá un valor teórico mínimo menor que cero que indicará asociación perfecta, pero en sentido inverso.
- Los estadísticos de asociación que cumplen con estas propiedades varían entre 0 y 1; o entre -1 y 1.

De esta manera, el coeficiente de asociación es un número que oscilará entre 0 y 1 o entre -1 y 1 y nos dice la probabilidad

que tenemos de intuir mejor los casos de una variable (*dependiente*) basándonos en el conocimiento que tenemos de una segunda variable (*independiente*). Así, en el caso de que cuando el número de casos en una dirección creciente aumente en ambas variables, diremos que nos encontramos ante una *asociación directa o positiva*. Si, en cambio, los incrementos se dan en direcciones opuestas, hablamos de una *asociación inversa o negativa* (Pérez Juste, 2012, pp. 42 y ss.).

Sin embargo, existen multitud de estadísticos de correlación que aportan información diferente en función del nivel de medición en el que se basen nuestras variables (Manheim y Rich, 2002). Y es que no en todos los casos se pueden usar los mismos estadísticos, dado que el nivel de medición utilizado con cada variable puede hacer que los resultados de unos u otros no sean interpretables. Para ello, se ha elaborado esta tabla 2 que resume de forma somera el coeficiente de asociación que tendríamos que utilizar en función del nivel de medición de nuestras variables, así como la información que nos ofrece.

Tabla 2. Coeficientes de asociación en función de niveles de medición

| Nivel de medición | Posibilidades | | Coeficientes |
|---------------------------------|--------------------------------|----------------------------------|-----------------------------------|
| Ambas dicotómicas | Existencia | | Chi-cuadrado |
| | Existencia y Magnitud | | V de Cramer |
| | Existencia, Magnitud y Sentido | | Phi Q de Yule |
| Ambas nominales | Pluricotómicas | Existencia | Chi-cuadrado |
| | | Existencia y Magnitud | V de Cramer |
| Ambas ordinales | Varias categorías | Existencia, Magnitud y Dirección | Rho de Spearman Tau-C Gamma |
| Una nominal – Otra intervalo | Existencia y Magnitud | | Etha |

Fuente: elaboración propia.

Dos de los coeficientes o estadísticos más habituales para expresar correlación son el Chi-cuadrado (χ^2) y la R de Pearson. Ambas pruebas estadísticas, ideadas por el estadístico inglés Karl Pearson (1857-1946) buscan comprobar la existencia de rela-

ción entre dos variables; es decir, comprobar que las variables no son independientes. Por un lado, Chi-cuadrado nos ofrece un valor estadístico cuyo número no tiene significado alguno, sino que nos interesa comprender si existe significatividad estadística en la relación. En caso positivo, diremos que existen relación entre nuestras variables. Por el otro lado, la R de Pearson indica, además de la existencia de relación entre las variables analizadas de la misma manera que Chi-cuadrado, el tipo de relación que se da. Esto es, su valor oscila entre -1 y 1 , significando que cuanto más cerca se encuentre el valor del -1 , que la relación es *indirecta* (a menos de una variable, más de la otra); si, por el contrario, el valor se acerca a 1 , diremos que la relación es *directa* (cuando una de las variables crece, la otra también, y viceversa). Así, aparte de decirnos si existe relación entre nuestras variables, indica la dirección y la fortaleza de esta (Pignataro, 2016).

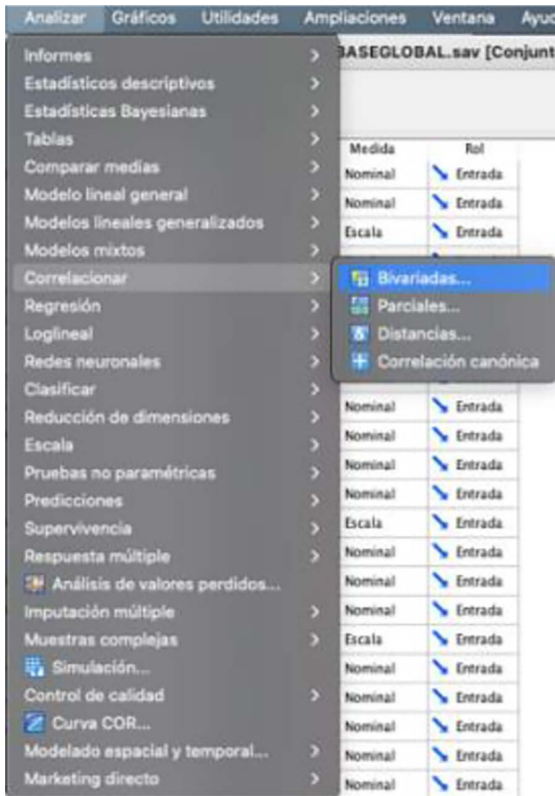


Figura 26. Proceso para solicitar correlación en SPSS. Fuente: elaboración propia.

Para poder realizar un análisis de correlación, existen diferentes opciones dentro de SPSS. Vamos a explicar aquí el más sencillo y habitual; existen otros que se mencionarán, pero que no se trabajarán con ejemplos visuales. Para poder comenzar a realizar una correlación, y tal y como se muestra en la figura 26, debemos seleccionar en el menú Analizar, la categoría Correlacionar y, en el menú desplegable, la opción Bivariadas. Una vez que lo hayamos seleccionado, se nos abrirá el cuadro de diálogo de la figura 27.

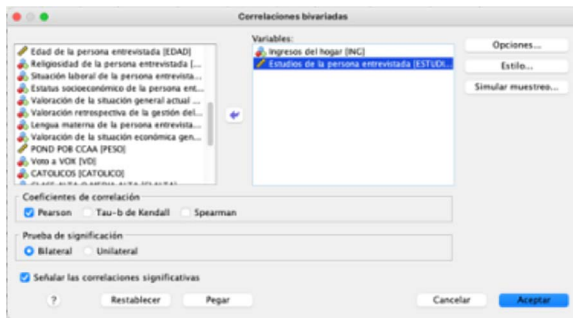


Figura 27. Cuadro de diálogo de correlaciones. Fuente: elaboración propia.

Una vez que tenemos abierto el cuadro de diálogo mostrado en la figura 27, debemos seleccionar las variables que queremos correlacionar de nuestra lista de variables que aparece a la izquierda como ya es habitual, y moverlas, ya sea arrastrándolas o a través del botón con la flecha a la sección central, en el cuadro con el rótulo «Variables». Debajo, en función de la tipología de nuestras variables, optaremos por aquel coeficiente que sea el más adecuado. Otro tipo de coeficientes más concretos pueden encontrarse en la opción descrita después de la figura 25, desde el cuadro de diálogo de tablas cruzadas que se muestra en la figura 21.

Para nuestro análisis, y al tratarse de dos variables que actúan como variables continuas, se ha optado por solicitar un coeficiente de correlación de Pearson. Este coeficiente fluctúa, tal y como se ha descrito en las páginas anteriores, entre -1 y 1 , siendo el primer valor una relación perfectamente negativa o inversa¹⁹ y,

19. Lo cual indica que cuando una variable aumenta en valor, la otra disminuye y viceversa.

el segundo, una relación perfectamente positiva o directa.²⁰ Si el valor está cercano a 0, indica que nuestras variables no tienen relación. Cabe destacar que, a través de este sistema de obtención de correlaciones, SPSS no identifica variables dependientes de variables independientes y procede a realizar el análisis de correlación para todas y cada una de las variables que se incluyan en la sección central del cuadro de diálogo de la figura 27.

| Correlaciones | | | |
|-------------------------------------|------------------------|--------------------|-------------------------------------|
| | | Ingresos del hogar | Estudios de la persona entrevistada |
| Ingresos del hogar | Correlación de Pearson | 1 | ,074** |
| | Sig. (bilateral) | | ,000 |
| | N | 17411 | 17107 |
| Estudios de la persona entrevistada | Correlación de Pearson | ,074** | 1 |
| | Sig. (bilateral) | ,000 | |
| | N | 17107 | 21909 |

** . La correlación es significativa en el nivel 0,01 (bilateral).

Figura 28. Resultado de correlaciones. Fuente: elaboración propia.

Una vez que hemos mandado a SPSS ejecutar el análisis de correlación, nos ofrecerá una tabla como la que aparece en la figura 28. En ella debemos destacar que, además del valor de correlación, aparece el de significación bilateral, lo cual nos indicará el nivel de confianza con el que estamos trabajando con nuestras variables.

Lo primero en lo que la atención recae es en el hecho de que la tabla muestra los resultados por duplicado. Y es que, como ya hemos calculado, SPSS realiza el análisis para los dos pares de variables. Es decir, en este caso ha calculado los estadísticos de correlación para el vínculo entre las variables ingresos del hogar y estudios, por un lado, y por otro, para las mismas variables de forma inversa. De ahí que todo se encuentre por duplicado.

Con relación a los estadísticos obtenidos, vemos cómo el valor de la correlación entre ingresos y estudios es de 0,074. Esto

20. Significando esto que, cuando una de las variables aumenta, la otra también lo hace, y viceversa.

provoca los siguientes supuestos. En primer lugar, el resultado es positivo lo cual indica una relación directa entre las variables; o sea, a mayores ingresos mayores estudios o, al revés, a mayor nivel de estudios, los ingresos del hogar aumentan. Sin embargo, el hecho de que el valor se encuentre muy cercano a cero dibuja un efecto pequeño. El segundo elemento que debemos tener en cuenta es el de la significación. En este caso el valor es de 0,000. Esto significa que nuestro p-valor es menor que 0,001, por lo que la probabilidad de que esta relación se dé es muy elevada. En otras palabras, estamos trabajando con una significatividad del 99 % y, por lo tanto, la relación entre nuestras variables es estadísticamente fuerte a pesar de tener un efecto muy pequeño, debido al valor de la R de Pearson.

Otra forma de solicitar correlaciones (y que incluye la práctica totalidad de coeficientes de asociación que se mostraban en la tabla 1) puede realizarse desde el menú de Tablas cruzadas que ya hemos comentado en la figura 21. En esa ventana y siempre después de haber incluido en las filas y columnas las variables que queremos analizar, haremos clic en el botón Estadísticos de la sección derecha (el segundo botón). De ahí, se abrirá el cuadro de diálogo que aparece representado en la figura 29.

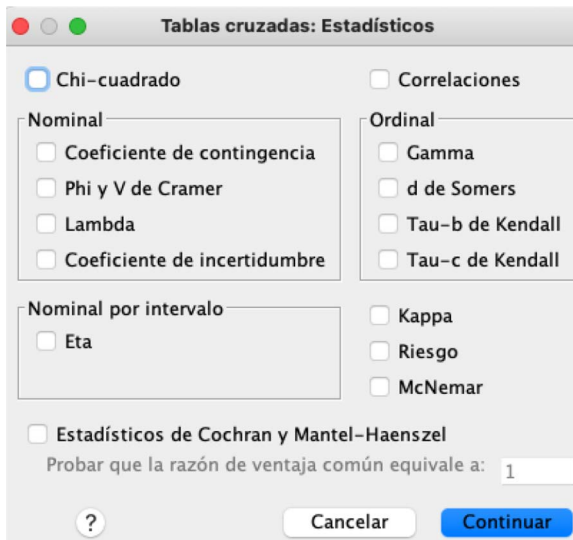


Figura 29. Cuadro de diálogo «Estadísticos». Fuente: elaboración propia.

Como se puede apreciar, en el cuadro de diálogo del menú Estadísticos de Tablas cruzadas, se nos da la posibilidad de obtener la Chi-cuadrado, las correlaciones (R de Pearson), y otra multitud de coeficientes de asociación agrupados por el nivel de medición de nuestras variables. Así, se puede comprobar en la figura 29 que se encuentran presentes en las diferentes agrupaciones los coeficientes de asociación mencionados en la tabla 1 y a los que deberemos acudir en función de las variables que queramos cruzar.

En caso de que nuestras variables sean de niveles de medición diferenciados, gracias a este cuadro de diálogo podremos marcar todos los coeficientes que vayamos a necesitar. Eso sí, posteriormente, tendremos que descartar los resultados de aquellas opciones que no resulten válidas en función de la tipología de variables en cada cruce.

Una vez que hemos seleccionado el coeficiente o los coeficientes que sean de interés para nuestro análisis, procederemos a clicar en el botón Continuar y en el de aceptar de las tablas cruzadas. De esta manera, y siempre que hayamos cubierto los pasos explicitados en las figuras 23 y 25, podemos ahorrar tiempo a la hora de realizar nuestro análisis bivariado a través de SPSS.

2.5.3. Presentación de resultados

Además del procedimiento estadístico, se han de saber presentar nuestros datos de una manera clara y ordenada. Gracias a lo anteriormente expuesto, la elección del gráfico adecuado nos ayudará a una correcta visualización de los resultados que queremos mostrar. También resulta clave la presentación de las tablas tanto para su comprensión como para su interpretación y seguimiento. Uno de los momentos clave es presentar los resultados de forma clara, concisa y ordenada. Es importante huir de un lenguaje farragoso o barroco (Kastellec y Leoni, 2007). Asimismo, se debe incluir siempre descripciones de los datos utilizados y analizados y, por supuesto, debe explicitarse si nuestra o nuestras hipótesis han sido corroboradas.

Para la exposición de datos en tablas, estas deben resumir y ordenar la información de nuestros análisis. Es muy importante la inclusión de los estadísticos más relevantes para nuestra investigación, y no únicamente las variables descriptivas. A la hora de

decidir si se utilizará una gráfica o una tabla, es importante atender a tres aspectos: la naturaleza de los datos a representar, los objetivos o propósitos de la tabla o gráfico, y el tipo de audiencia al que se dirige el texto (Garza Mercado, 1967).

En las figuras 30 y 31 se plasman dos ejemplos sobre cómo presentar una tabla y un gráfico en una investigación, incluyendo todos los elementos relevantes. En el caso de la figura 30, mostramos cómo se estructuraría una tabla con un total de cuatro variables representadas. En la tabla se aprecia el título y la fuente, dos elementos fundamentales, así como las diferentes categorías de las variables que se van a analizar (los países) y los datos referentes a cada país para cada una de las variables de análisis. Esta tabla se ha construido a través de la agregación de varias tablas cruzadas como las que se han visto en el anterior epígrafe, tomando siempre la variable país como dependiente y el resto como independientes.

| Tabla 1. Resumen de la presencia femenina en las cámaras | | | |
|--|------------------|----------------|----------------------|
| País | Mínimo por lista | Presencia real | Mínimo vs. Presencia |
| Argentina | 30% | 40,86% | Más |
| Bolivia | 50% | 53,08% | Más |
| Brasil | 30% | 15,01% | Menos |
| Chile | 40% | 22,58% | Menos |
| Colombia | 30% | 18,71% | Menos |
| Ecuador | 50% | 37,96% | Menos |
| Paraguay | 20% | 15% | Menos |
| Perú | 30% | 30% | Igual |
| Uruguay | 33% | 19,19% | Menos |
| Venezuela | | 22,16% | |
| Fuente: elaboración propia | | | |

Figura 30. Ejemplo de presentación de una tabla en una investigación. Fuente: García Santamaría y Pérez Castaños (2020, p. 825).

En cuanto a la presentación de los datos en un gráfico, la figura 31 muestra un ejemplo de la misma publicación que la tabla. De nuevo, se observa cómo se dispone de dos elementos fundamentales: el título y la fuente. Es muy importante que el título sea conciso y condense lo que se va a mostrar, y la fuente tiene que indicar de dónde hemos obtenido nuestros datos.

Además, se ha optado por un gráfico de líneas múltiples en donde se ofrece la evolución de la variable objeto de análisis en tres países de manera longitudinal, comprobando sus cambios con el paso del tiempo para cada variable (país) y permitiendo su comparación con el resto de las variables también de manera longitudinal.

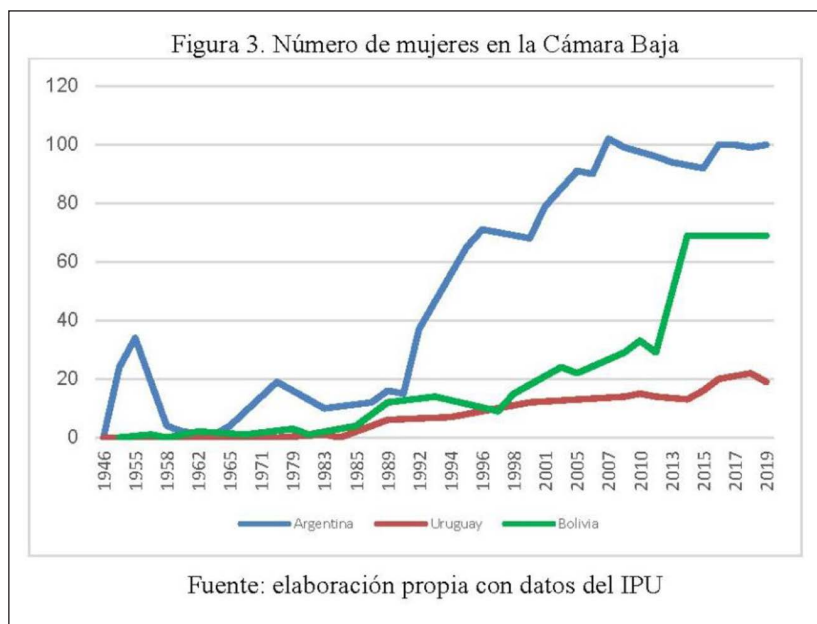


Figura 31. Ejemplo de presentación de un gráfico en una investigación. Fuente: García Santamaría y Pérez Castaños (2020, p. 822).

En cualquiera de los casos, es clave guiar al lector a través de los resultados, asegurándose de que quedan todos los capitales reflejados. Además, ha de referenciarse si las guías teóricas utilizadas se corresponden con nuestros datos o no. Primero, partiremos de lo más sencillo (frecuencias) e iremos avanzando hacia las partes más complejas. Nunca deben presentarse todos los datos, sino solo aquellos que resuman las partes más importantes de nuestra investigación. Finalmente, hay que recordar que tablas y gráficos deben ir siempre numerados, generalmente en caracteres arábigos, seguidos de un breve título que resuma su contenido, y siempre con indicación de la fuente de obtención de los datos.

Sobre la presentación de las tablas, debe destacarse que es importante incluir las variables que se están analizando, así como el dato que se ofrece, indicando esto en el encabezado (N, Porcentaje, variable...). Es conveniente también que evitemos al máximo las abreviaturas, aunque algunas pueden utilizarse. Al pie de la tabla se incluye la fuente, así como las explicaciones necesarias, siempre que las hubiere, las cuales, además, deben referenciarse en la misma tabla.

2.6. Fiabilidad y validez de cuestionarios y escalas

Comprobar hasta qué punto lo que tratamos de medir se ha realizado de forma correcta y, por lo tanto, cumple con unas condiciones mínimas de validez y fiabilidad resulta fundamental. Y es que, además de los problemas en la operacionalización ya descritos en el apartado anterior, debemos destacar la existencia de dos tipos de errores (Casas-Casas, 2013):

- *Errores sistemáticos*: que son aquellos que producen sesgo y, por ello, ponen en cuestión los resultados de la investigación.
- *Errores aleatorios*: que son menos graves, en la medida en que se compensan unos a otros. Sin embargo, debilita la precisión en la medición del concepto.

La fiabilidad y la validez deben ser propiedades de los instrumentos de medición para evitar tanto errores sistemáticos como aleatorios. Así, la validez se corresponde con el grado en que los indicadores miden los conceptos que se quieren medir. Por ello, es el grado en el que la variación de un indicador refleja únicamente diferencias en el concepto que se intenta medir (Coronado, 2007). No se puede lograr una validez completa, ya que en el proceso de operacionalización los conceptos se empobrecen. La validez está relacionada con el error sistemático. Un indicador válido no produce errores sistemáticos.

Por su parte, la fiabilidad es el grado de precisión del instrumento de medida o indicador que mide el concepto. Cuando un

instrumento de medición proporciona distintos resultados al aplicarlo varias veces es poco fiable. También hay problemas de fiabilidad cuando en una clasificación es posible asignar un caso a más de una categoría. La fiabilidad está asociada al error no sistemático (Fernández-Esquinas, 2003).

2.6.1. Fiabilidad como consistencia interna

Como ya hemos indicado con anterioridad, la fiabilidad expresa la consistencia de una medida; es decir, cómo lo que queremos medir se puede reproducir o puede ser medido por otras personas que realicen investigaciones con objeto de estudio idéntico o similar.

El procedimiento más conocido es el denominado *test-retest*, que consiste en aplicar el mismo instrumento, en dos ocasiones distintas, a los mismos sujetos. Este procedimiento ayuda a medir la estabilidad del instrumento. Para poder comprobar la fiabilidad de un instrumento o de un cuestionario tenemos primero que asegurarnos que la característica o características que queremos medir no varíen con el transcurso del tiempo. Además, es preciso que pase un tiempo suficiente como para que la persona entrevistada no recuerde la primera respuesta que dio. La fiabilidad se obtiene con un análisis de correlación entre ambas mediciones (Gil, 2006). De esta manera, si obtenemos una correlación alta entre los resultados de las dos aplicaciones, dispondríamos de una buena fiabilidad del instrumento.

Una segunda forma de comprobar la fiabilidad es a través del método de versiones alternativas. Aquí lo que se busca es, a través de aplicar dos versiones diferentes de la misma herramienta a la misma muestra, obtener resultados similares. La cuestión central reside en lograr dos redacciones lo bastante diferentes y, sin embargo, equivalentes (Murillo *et al.*, 2014). Una vez recogidos los resultados se efectúa un análisis de fiabilidad medido a través del Alfa de Cronbach que comparará ambas versiones. Este estadístico, que fluctúa entre 0 y 1 va a indicar el grado de fortaleza que tienen las dos herramientas para medir, teóricamente, lo mismo. Cuanto más cercano se encuentre el valor a 1, mayor fiabilidad tendrá nuestra medida. Las próximas tres figuras explicarán cómo calcularlo a través de SPSS.

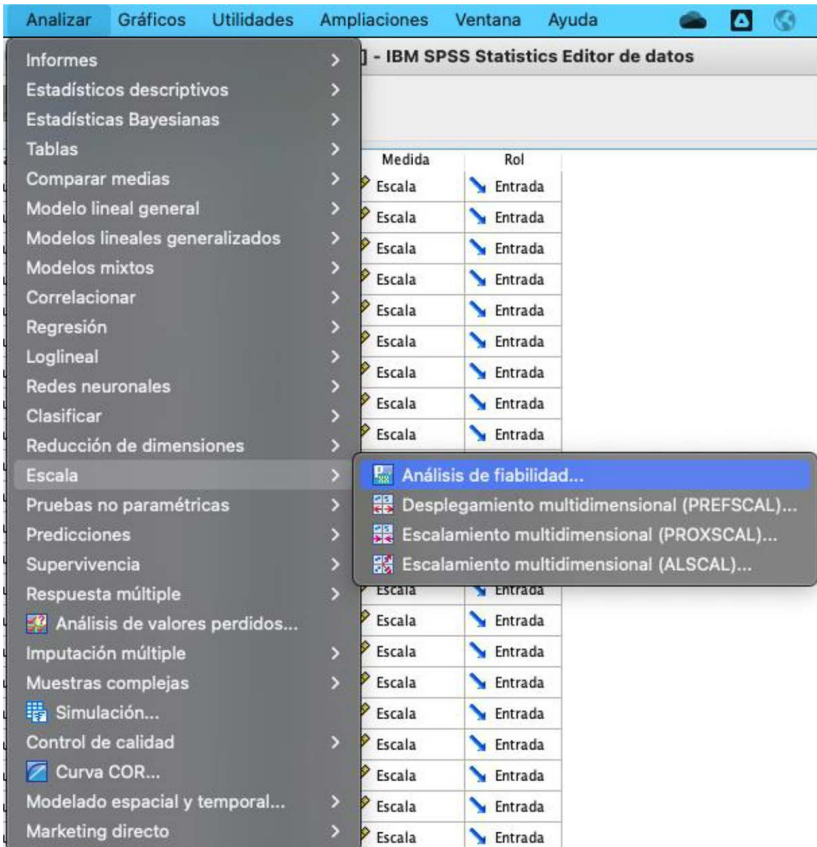


Figura 32. Proceso para solicitar análisis de fiabilidad. Fuente: elaboración propia.

El comando para solicitar el análisis de fiabilidad se encuentra, como todos los que previamente hemos descrito, en el menú *Análisis*. Dentro del menú desplegable, esta vez debemos ir a la parte media-baja del desplegable, hasta que encontremos la opción *Escala*. Allí, al posar nuestro cursor, se desplegará un nuevo menú con cuatro opciones, siendo la primera la que nos interesa: *Análisis de fiabilidad*.

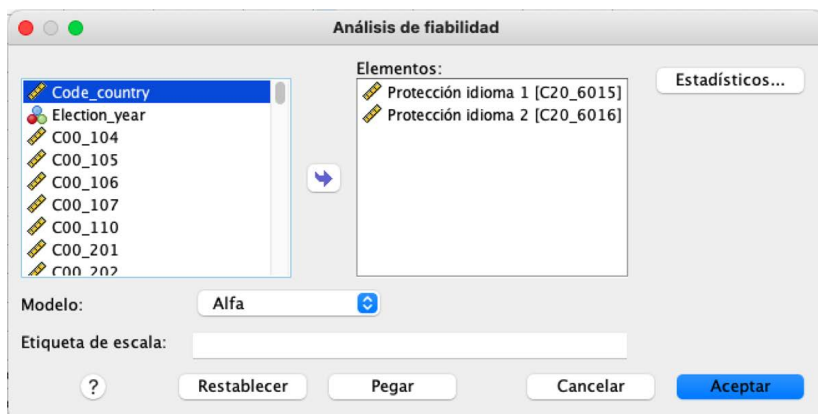


Figura 33. Cuadro de diálogo de la fiabilidad. Fuente: elaboración propia.

Una vez que hemos hecho clic en Análisis de fiabilidad, se nos abrirá el cuadro de diálogo que aparece en la figura 33. Aquí tenemos múltiples opciones, pero, a efectos prácticos, nos centraremos en lo que es más relevante a lo que se ha comentado. Al igual que sucedía en otros cuadros de diálogo, en la parte izquierda disponemos de un cuadro con todas nuestras variables que podremos incluir en el análisis. En la parte central aparece un cuadro en donde se incluirán aquellas variables que queremos testar. En este caso, hemos optado por dos variables que miden el nivel de protección de los idiomas no mayoritarios en los países. Son dos formas diferentes de obtener un índice que habla de cómo se protegen, fomentan o defienden los idiomas minoritarios.

Entre los dos cuadros anteriormente comentados, encontramos un pequeño cuadro en donde aparece la palabra Alfa. En él, a través del desplegable, podemos seleccionar una serie de test de acuerdo con el que más nos interese para nuestras variables. En este caso, se corresponde con el Alfa de Cronbach, por lo que no debemos cambiarlo. En la parte derecha, se encuentra únicamente un botón, el de Estadísticos, que nos permitirá obtener información adicional de nuestras variables como puede ser la frecuencia, medidas de tendencia central o correlaciones entre otras cuestiones. De nuevo, por cuestiones prácticas, no profundizaremos en el botón.

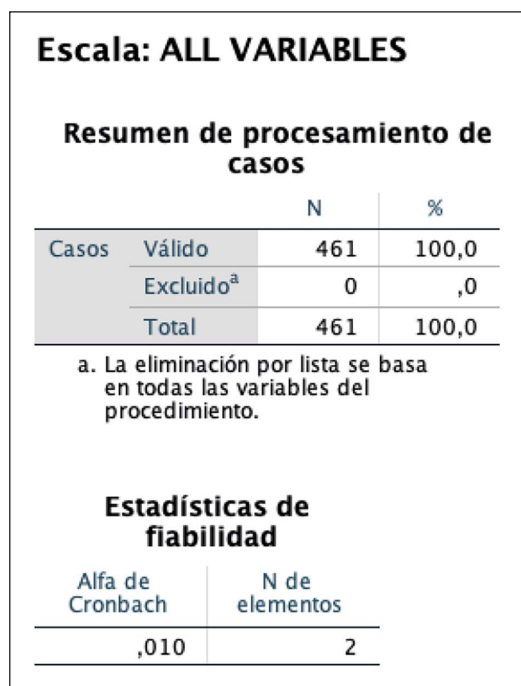


Figura 34. Resultados de la fiabilidad. Fuente: elaboración propia.

Una vez que hayamos seleccionado las variables de las que nos interesa obtener el estadístico de fiabilidad, procedemos a hacer clic en el botón Aceptar y el resultado que se plasmará en nuestra ventana de resultados de SPSS es el que vemos reproducido en la figura 34. El programa estadístico nos ofrece una primera tabla resumen de nuestras variables y nos indica que, para nuestro análisis, ha utilizado la totalidad de la base de datos, los 461 casos disponibles, no excluyendo ninguno.

Después, se presenta la tabla con los resultados estadísticos propiamente dichos. En este caso, la tabla consta de dos columnas, una primera con el valor del Alfa de Cronbach, y una segunda con el número de elementos que se han incluido en el análisis (dos en este caso). Como hemos explicado con anterioridad, el Alfa fluctúa entre 0 y 1. Cuanto más cerca de cero, menor fiabilidad y viceversa. En términos analíticos, en Ciencias Sociales los valores superiores a 0,7 nos indican que tenemos una fiabilidad elevada, pero, sobre todo, buscamos valores de 0,8 o más. En el ejemplo de la figura 34, vemos que el valor ofrecido por

SPSS es de 0,01. Esto significa que nuestras dos medidas no son fiables, o, lo que es lo mismo, que no miden el concepto que tratamos de medir de una forma adecuada.

Por último, volviendo a las formas que tenemos de comprobar la fiabilidad, encontramos el procedimiento de las dos mitades, que es menos habitual en la medida en que requiere dividir nuestras preguntas en dos mitades que busquen abordar un contenido similar. Es decir, a partir de la división de una prueba o herramienta en dos mitades, se obtienen dos puntuaciones para cada sujeto (Castro Posada, 2002). Por ejemplo, una escala de 12 ítems se divide por la mitad y cada una de ellas se aplica, como si estuviera completa, a un mismo grupo. Para estimar la fiabilidad calculamos la varianza de las diferencias entre las dos mitades y la varianza de la escala al completo. Así, una menor varianza indicará más homogeneidad de resultados y, por ende, mayor fiabilidad de nuestra herramienta.

Para finalizar, como señala Cea D'Ancona (1998:156), conviene insistir en que la falta de fiabilidad puede deberse desde a que nuestro instrumento de medida no es adecuado (preguntas mal redactadas en un cuestionario) hasta a cómo se aplica el instrumento (por ej.: el formato de la encuesta), pasando por cómo se analiza la información obtenida. Con todo ello, cabe destacar que, al menos de manera realista, alcanzar una fiabilidad perfecta es bastante difícil de alcanzar.

2.6.2. Validez de constructo

A la hora de abordar la validez encontramos diferentes cuestiones como las que vamos a ver a continuación. La primera de ellas es la validez de constructo y trata de demostrar que aquello que mide el instrumento es una variable consistente, comparando una medida particular con aquella que teóricamente se habría de esperar (a partir de las hipótesis derivadas del marco teórico de la investigación). Es decir, se refiere al grado de adecuación alcanzado en la medición de los conceptos centrales de la investigación.

Y es que, como los conceptos pueden medirse de distinta forma y es difícil que puedan contemplarse todas las dimensiones, se recomienda ser riguroso en la operacionalización y efectuar operacionalizaciones múltiples para asegurarnos que, efectiva-

mente, estamos midiendo lo que queremos medir. Para ello es recomendable buscar dos o más mediciones del concepto y comprobar en qué medida coinciden (Manheim y Rich, 2002).

Ello muestra clara la importancia de contar con indicadores múltiples para nuestras variables. Siempre que dispongamos de medidas múltiples, no solo nos permitirá verificar la validez de nuestros indicadores, sino que también mejorará nuestras posibilidades de obtener una medida válida de las variables, haciendo que mejore verdaderamente la validez. Y es que combinar los resultados de diversos procedimientos de medida para obtener una calificación compuesta será un reflejo válido del valor real de nuestra variable. Esto se conseguirá de forma conjunta con más probabilidad que cualquiera de las medidas tomadas por separado. Además, aumenta la posibilidad de que los errores que puedan invalidar cada medida queden eliminados al combinar varios procedimientos de medición (Fernández-Esquinas, 2003).

2.6.3. Validez de contenido

Cuando nos referimos a la validez del contenido en realidad nos estamos refiriendo a una combinación de tres factores diferenciados. Por un lado, hablamos de la *validez interna*, por otro, de la *validez externa* y, finalmente, de la *validez estadística*. Estos tres elementos combinados ayudan a alcanzar herramientas e indicadores de medición adecuados cuando estamos realizando análisis en Ciencias Sociales (Manheim y Rich, 2002).

Comenzando por la validez interna, estamos hablando de que nuestros indicadores o variables cumplan con la posibilidad de establecer relaciones de causalidad entre variables, controlando otras posibles explicaciones (Krippendorff, 2004). Y es que la validez interna viene garantizada por el control de explicaciones alternativas, pudiéndose hacer a priori o a posteriori. Habitualmente, en nuestras disciplinas, optamos por esta segunda opción, para la que nos valemos de técnicas estadísticas multivariantes.²¹ Es importante destacar que, si no cumplimos

21. Estas técnicas, que avanzan más allá del contenido de este volumen, pueden comprobarse en Manheim y Rich (2002) o en Krippendorff (2004) y se corresponderá con análisis factoriales o de regresión lineal o logística entre otros.

con la validez interna, toda nuestra investigación podría quedar en entredicho al no estar consiguiendo medir lo que efectivamente nos habíamos propuesto en un inicio, invalidando los resultados y conclusiones a los que podríamos llegar.

Por otro lado, como decíamos, nos encontramos con la validez externa. Esta se encuentra referida a la posibilidad de realizar generalizaciones a partir de los resultados obtenidos en nuestra investigación; es decir, a la población objeto de estudio (Coronado, 2007). Nos centramos en que la herramienta aplicada puede ser válida para nuestra muestra, pero ¿nos asegura que llega a medir lo que queremos medir para el conjunto de la población? Para ello, es fundamental que la muestra sea verdaderamente representativa del conjunto, tanto en número de casos como en el procedimiento de selección, por lo que se recomienda seguir procedimientos aleatorios o probabilísticos. Como se puede apreciar, en relación con la validez externa, el paso para asegurarnos su cumplimiento se da con anterioridad a la aplicación de cualquiera de las herramientas o instrumentos y no

Finalmente, la última de las características conformantes de la validez de contenido es la que se refiere a la validez estadística. Esta pone en diálogo la fiabilidad anteriormente comentada con el análisis estadístico y es que en lo que pone el foco es en la adecuación y fiabilidad de la técnica de análisis de datos seleccionada, condicionado por el tamaño muestral y el cumplimiento de una serie de criterios (normalidad, homocedasticidad, etc.), todos replicables a través, de nuevo, de técnicas multivariantes (Lynch, 2013).

Además, Krippendorff (2004) estructura la validez a través de su relevancia nominal, social y empírica. La primera de ellas, la nominal, se basa en la obtención de conclusiones razonables, que tengan sentido; la segunda, la social, tiene su fundamentación en el grado de importancia y significado del análisis para la sociedad; y, finalmente, la validez empírica se establece en el grado en el que los datos disponibles y la teoría establecida confirman varias etapas en un proceso de investigación, el grado en el que las inferencias específicas resisten el reto de datos adicionales, de las conclusiones de otros trabajos de investigación.

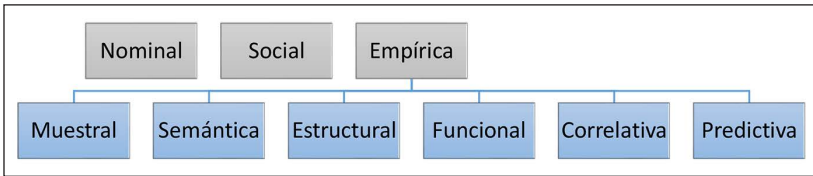


Figura 35. Diagrama de la validez empírica. Fuente: elaboración propia a partir de Krippendorff (2004).

Tal y como se muestra en la figura 35, la validez empírica se divide a su vez en seis subcategorías diferenciadas. Por su parte, la validez muestral establece si la muestra que se utiliza en la investigación es representativa de la población. La validez semántica se consigue cuando las categorías utilizadas tienen razón de ser en el contexto en el que se crean. La tercera es la validez estructural, que se establece cuando los datos y la teoría se corresponden con el análisis empleado. La cuarta de las subdivisiones es la validez funcional, que busca la existencia de correspondencia entre nuestros resultados y lo ya investigado. La validez correlativa denota el hecho de que, a mayor correlación, mayor será la validez de los datos utilizados y de la investigación en general. Finalmente, la sexta es la validez predictiva que se antoja en una de las principales en las Ciencias Sociales. Esta no se alcanza de otra manera que no sea a través de la capacidad de nuestros análisis e investigaciones para poder predecir acontecimientos futuros y, en función de estos, prevenirlos, anticiparnos o fomentarlos.

2.7. Referencias bibliográficas

- Alaminos, A. (1993). *Gráficos*. CIS
- Alvira, F. y Martínez Ramos, E. (1985). El efecto de los entrevistadores sobre las respuestas de los entrevistados. *Revista Española de Investigaciones Sociológicas*, 29, 219-256.
- Anduiza, E., Crespo, I. y Méndez, M. (2011). *Metodología de la Ciencia Política*. CIS.
- Asimov, I. (1979). *Introducción a la Ciencia*. Plaza y Janés.
- Batthyány, K. y Cabrera, M. (2011). *Metodología de la Investigación en Ciencias Sociales. Apuntes para un curso inicial*. Universidad de la República.

- Beltrán, M. (2000). Cinco vías de acceso a la realidad social. En: García Fernando, M., Ibáñez, J. y Alvira, F. *El análisis de la realidad social. Métodos y técnicas de investigación social* (pp. 15-55). Alianza.
- Bonilla Castro, E. y Rodríguez Sehk, P. (1997). *Más allá del dilema de los métodos. La investigación en ciencias sociales*. Uniandes.
- Bourque, L. B. y Clark, V. A. (2008). *Processing Data: The Survey Example*. Sage.
- Camel, F. (1970). *Estadística médica y de salud pública*. Universidad de los Andes
- Campbell, D. T. y Stanley, J. C. (2002). *Diseños experimentales y cuasiexperimentales en la investigación social*. Amorrortu.
- Carrasco Arroyo, S. (2012). *Aproximación a la estadística desde las ciencias sociales*. Universidad de Valencia.
- Casas-Casas, A. (2013). El uso de métodos experimentales en ciencia política. En: Casas-Casas, A. y Méndez Méndez, N. *Experimentos en ciencias sociales: usos métodos y aplicaciones* (pp. 53-74). PUJ.
- Castro Posada, J. A. (2001). *Metodología de la investigación: Fundamentos*. Amarú.
- Castro Posada, J. A. (2002). *Metodología de la investigación: Diseños*. Amarú.
- Cea D'Ancona, M. A. (1998). *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Síntesis.
- Cicourel, A. V. (2011). *Método y medida en sociología*. CIS.
- Corbetta, P. (2007). *Metodología y técnicas de investigación social*. McGraw Hill.
- Coronado, J. (2007). Escalas de medición. *Paradigmas*, 2(2), 104-125.
- Cortés, F. y Rubalcava, R. M. (1987). *Métodos estadísticos aplicados a la investigación en ciencias sociales: análisis de asociación*. Colegio de México.
- Crespo, I. (2016). Fases de una investigación aplicada mediante encuesta I. El planteamiento del problema y II. El diseño de la investigación. En: Crespo, I. et al. (eds.). *Manual de herramientas para la investigación de la opinión pública* (pp. 90-102) Tirant lo Blanch.
- De Leeuw, E. (2004). *New technologies in data collection, questionnaire design and quality*. Instituto Vasco de Estadística.
- Denzin, N. K. y Lincoln, Y. S. (2005). *The Sage Handbook of Qualitative Research*. Sage.
- Díaz de Rada, V. (2001). *Diseño y elaboración de cuestionarios para la investigación comercial*. ESIC.
- Díaz de Rada, V. (2002). *Tipos de encuestas y diseños de investigación*. Universidad de Navarra.

- Díaz de Rada, V. (2005). *Manual de trabajo de campo en la encuesta*. CIS.
- Dillman, D., Smyth, J. y Christian, L. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley.
- Emmerich, G. E. (1997). *Metodología de la Ciencia Política*. UAM-I.
- Fernández-Esquinas, M. (2003). Criterios de calidad en la investigación social: la producción de datos sociales. *Empiria*, 6, 47-77.
- Font Fàbregas, J. y Pasadas del Amo, S. (2016). *Las encuestas de opinión*. CSIC.
- Fowler, F. J. (1995). *Improving Survey Questions*. Sage.
- Frías-Navarro, M. D. (2008). *Evaluación empírica de las hipótesis*. Universidad de Valencia.
- García, E. et al. (2011). *Inferencia estadística*. Garceta.
- García Ferrando, M. (1993). La encuesta. En: García, M., Ibáñez, J. y Alvira, F. *El análisis de la realidad social. Métodos y técnicas de Investigación* (pp. 245-283). Alianza.
- García Santamaría, S. y Pérez Castaños, S. (2020). Representación parlamentaria femenina en Latinoamérica. Del dicho al hecho. *Revista Opción*, 93(2), 796-833.
- Garza Mercado, A. (1967). *Manual de técnicas para estudiantes de ciencias sociales*. Universidad de Nuevo León.
- Gil, P. (2006). *Estadística e Informática (SPSS) en la investigación descriptiva e inferencial*. UNED.
- Giner, S. (1983). *Sociología*. Península.
- Hernández Sampieri, R., Fernández-Collado, C. y Baptista Lucio, P. (2010). *Metodología de la investigación*. McGraw-Hill.
- Kaltenbach, H. M. (2012). *A Concise Guide to Statistics*. Springer.
- Kastellec, J. P. y Leoni, E. L. (2007). Using Graphs instead of tables in Political Science. *Perspectives on Politics*, 5(4), 755-771.
- Krippendorff, K. (2004). Reliability in Content Analysis. *Human Communication Research*, 30(3), 411-433.
- Krueger, R. A. (1991). *El grupo de discusión. Guía práctica para la investigación aplicada*. Pirámide.
- Lavrakas, P. J. (1993). *Telephone Survey Methods*. Sage.
- Lazarsfeld, P. y Oberschall, A. R. (1965). Max Weber and Empirical Social Research. *American Sociological Review*, 30(2), 185-198.
- López-Roldán, P. y Fachelli, S. (2015). *Metodología de la investigación social cuantitativa*. Universidad Autónoma de Barcelona.
- Lynch, S. M. (2013). *Using Statistics in Social Research. A concise approach*. Springer.

- Manheim, J. B. y Rich, R. C. (2002). *Análisis político empírico. Métodos de investigación en ciencia política*. Alianza.
- Martínez, V. (2016). Fases de una investigación aplicada mediante encuesta IV. La elaboración de la muestra. En: Crespo, I. et al. (eds.). *Manual de herramientas para la investigación de la opinión pública*. Tirant lo Blanch.
- Mayntz, R., Holm, K. y Hübner, P. (1975). *Introducción a los métodos de la sociología empírica*. Alianza.
- Minujin, A., D'Alessio A. y Oiberman, I. (1987). *Estadística Descriptiva. Manual orientado a las Ciencias sociales*. Punto Sur.
- Moreno, C. y Mora, A. (2016). La conceptualización empírica de la opinión pública. En: Crespo, I. et al. (eds.). *Manual de herramientas para la investigación de la opinión pública*. Tirant lo Blanch.
- Murillo, F. J., Hernández-Castilla, R., Hidalgo, N. y Martínez-Garrido, C. (2014). Elaboración y Evaluación Psicométrica de la Escala de Actitudes hacia la Justicia Social en Educación (EAJSE). *Revista Internacional de Educación para la Justicia Social*, 3(2), 215-233.
- Newell, R. (1993). Questionnaires. En: Gilbert, N. (comp.). *Researching Social life*. Sage.
- Osgood, C., Suci, G. y Tannenbaum, P. (1957). *The measurement of meaning*. University of Illinois Press.
- Pasadas, S. (2014). *Población «solo-móvil» y precisión de las encuestas prelectorales basadas en el modo de administración telefónico. El caso de las elecciones andaluzas de 2012* (tesis doctoral). Departamento de Sociología, Universidad Pública de Navarra, Pamplona, España.
- Pérez Castaños, S. y Trujillo Cerezo, J. M. (2018). La encuesta como técnica de investigación en Criminología. En: Grimaldo Santamaría, R. O. (coord.). *Metodología de la Investigación en Criminología: Aspectos teóricos y prácticos* (pp. 59-95). Delta.
- Pérez Juste, R. (2012). *Estadística aplicada a las Ciencias Sociales*. UNED.
- Pignataro, A. (2016). *Manual para el análisis político cuantitativo*. UCR.
- Polit, D. y Hungler, B. (1987). *Investigación científica en ciencias de la salud*. McGraw Hill.
- Price, J. H. y Murnan, J. (2004). Research Limitations and the Necessity of Reporting Them. *American Journal of Health Education*, 35, 66-67.
- Robles Sastre, E. (2002). *Metodología e Investigación. Contenidos y Fases*. UCJC.
- Rodríguez, G., Gil, J. y García, E. (1996). *Metodología de la investigación cualitativa*. Aljibe.
- Rodríguez-Osuna, J. (2001). *Métodos de muestreo*. CIS.

- Romo, J. (2014). *Introducción a la estadística para las Ciencias Sociales*. McGraw Hill.
- Rossi, P. (1982). La sociología nella seconda metà del'ottocento: Dall'impiego di schemi storico-evolutivi alla formulazione di modelli analitici. *Il pensiero político*, 15(1), 188-215.
- Rubio, M. J. y Varas, J. (1997). *El análisis de la realidad en la intervención social. Métodos y técnicas de investigación*. CCS.
- Samaja J. (2003). *Epistemología y metodología. Elementos para una teoría de la investigación científica*. Eudeba.
- Sheatsley, P. (1983). Questionnaire Construction and Item Writing. En: Rossi, P., Wright, J. y Anderson, A. (eds.). *Handbook of Survey Research*. Academic Press.
- Sierra Bravo, R. (1994). *Técnicas de investigación social*. Paraninfo.
- Tomeo, V. y Uña, I. (2009). *Estadística descriptiva*. Garceta.
- Vallés Martínez, M. S. (2000). *Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional*. Síntesis.
- Weber, M. (2012). *Ensayos sobre metodología sociológica*. Amorrortu.
- Wonnacott, T. H. y Wonnacott, R. J. (2004). *Introducción a la Estadística*. Limusa.