

Cecilia Blanco



Encuesta y Estadística

Métodos de Investigación Cuantitativa
en Ciencias Sociales y Comunicación

ENCUESTA Y ESTADÍSTICA

MÉTODOS DE INVESTIGACIÓN CUANTITATIVA EN CIENCIAS
SOCIALES Y COMUNICACIÓN

Cecilia Blanco

ENCUESTA Y ESTADÍSTICA

MÉTODOS DE INVESTIGACIÓN CUANTITATIVA
EN CIENCIAS SOCIALES Y COMUNICACIÓN

 Editorial Brujas

Blanco, Cecilia

Encuestas y estadísticas : métodos de investigación cuantitativa en ciencias sociales y comunicación . - 1a ed. - Córdoba : Brujas, 2011.
160 p. ; 24x15 cm.

ISBN 978-987-591-276-2

1. Estadísticas. I. Título
CDD 310.4

© 2011 Editorial Brujas

1° Edición.

Impreso en Argentina

ISBN: 978-987-591-276-2

Queda hecho el depósito que marca la ley 11.723.

Ninguna parte de esta publicación, incluido el diseño de tapa, puede ser reproducida, almacenada o transmitida por ningún medio, ya sea electrónico, químico, mecánico, óptico, de grabación o por fotocopia sin autorización previa.



www.editorialbrujas.com.ar publicaciones@editorialbrujas.com.ar

Tel/fax: (0351) 4606044 / 4691616- Pasaje España 1485 Córdoba - Argentina.

AGRADECIMIENTOS

Al Lic. Prof. Julio Carballo, quien en sus clases de Estadística del 4º año de la Licenciatura en Comunicación Social, de la Escuela de Ciencias de la Información (ECI) de la Universidad Nacional de Córdoba (UNC) me transmitiera la pasión por la investigación cuantitativa.

A la Prof. Contadora Graciela Rodríguez, quien me permitiera profundizar sobre los fundamentos de la Estadística Descriptiva.

A mi querida Prof. y Estadística Nidia Blanch, quien además de formarme rigurosamente en el marco del Posgrado en Estadística Aplicada a la Investigación, de la Facultad de Ciencias Económicas (FCE) de la UNC, ha evaluado y validado los procedimientos y fórmulas estadística/os presentes en este libro

A la Dra. Paulina Emanuelli, por haberme permitido realizar una investigación metodológica en relación con el método de encuesta en el marco de la materia Metodología de la Investigación Aplicada a la Comunicación Social, de la ECI, de la UNC.

A la Lic. Gladys Ambroggio, de la Escuela de Ciencias de la Educación (ECE), de la Facultad de Filosofía y Humanidades (FFyH) de la UNC, por haberme formado en el manejo y tratamiento de grandes bases de datos y en el análisis e interpretación de datos estadísticos.

Al Prof. Dr. Eduardo Bologna, por su dedicada lectura de esta obra y sus invaluable aportes para su superación

Sin todos ellos, este proyecto no hubiera sido posible.

Índice

PRESENTACIÓN.....	13
-------------------	----

CAPÍTULO I

1. UNIVERSO, VARIABLES Y MUESTREO	23
1.1 Estadística y pensamiento estadístico.....	23
1.2. Unidades de análisis, variables, valor y dato.....	24
1.3 Tipos de variables.....	25
1.4 Grados de medición de las variables	27
1.5 Operacionalización de variables.....	31
1.5.1 Procedimiento de operacionalización de variables ...	31
1.5.2 Ejemplo de operacionalización de la variable “clase social” ...	34
1.5.3 La operacionalización de variables en el contexto de una investigación científica. Su implicancia en el proceso de investigación y el armado de cuestionarios. Un ejemplo desde la Teoría de la acción social de Pierre Bourdieu.	35
1.5.4 Ejemplo de operacionalización de variables a partir de la Teoría de usos y gratificaciones. Implicancias para la construcción de un cuestionario.	39
1.5.5 Operacionalización de variables en el contexto empresario/institucional	42
1.6 Población y muestra	43
1.7 Muestreo.....	44
1.7.1 Población, unidades de muestreo, extensión y tiempo..	44
1.7.2 Marco de muestreo.....	45
1.7.3 Definición del tamaño muestral	45
1.7.4 Elección del método de selección de muestras	55

CAPÍTULO II

2. TÉCNICA DE ENCUESTA: EL CUESTIONARIO Y SU APLICACIÓN.....	71
2.1 Un poco de historia.....	71
2.2 Algunos supuestos de la técnica de encuesta.....	73
2.3 Tipos de estudios.....	74
2.3.1 Estudios comparados.....	74
2.3.2 Estudios longitudinales.....	74
2.4 ¿Qué es y para qué sirve la encuesta?	75
2.5 Cuestiones éticas.....	75
2.6 Cuestiones generales de procedimiento	76
2.7 Construcción del cuestionario.....	76
2.7.1 Determinación de variables, nivel de medición y operacionalización.....	76
2.7.2 Tipo, cantidad, disposición de las preguntas y su codificación	77
2.8 Prueba piloto del instrumento.....	78
2.9 Recomendaciones para armar un cuestionario.....	79
2.9.1 Presentación del cuestionario.....	79
2.9.2 Recomendaciones para redactar buenas preguntas ..	81
2.9.3 Tipo de preguntas.....	86
2.10 Formas de administración del cuestionario.....	99
2.11 Procesamiento de la información: numeración de cédulas y codificación.....	101
2.12 Codificación	102
2.12.1 Codificación de preguntas y categorías:	102
2.13 Libro de códigos.....	103
2.14 Codificación física.....	104
2.15 Matriz de datos	105
2.16 Codificación manual.....	105

CAPÍTULO III

3. RUDIMENTOS DE ESTADÍSTICA DESCRIPTIVA	107
3.1 Análisis y lectura de datos	107
3.1.1 Tabla de frecuencias.....	107

3.1.2 Medidas de tendencia central	115
3.1.3 Medidas de dispersión	134
3.2 Algunas estrategias de redacción de informes estadísticos	140
3.2.1 Un ejemplo de lectura de datos.....	141
3.3 Ejercicios.....	142
Bibliografía	157

PRESENTACIÓN

El presente trabajo es el resultado de una investigación metodológica llevada a cabo en el marco de la materia Metodología de la Investigación aplicada a la Comunicación Social, de la Escuela de Ciencias de la Información, de la Universidad Nacional de Córdoba. Al mismo tiempo, obedece a la propia experiencia como investigadora dentro de algunos equipos de investigación a los que hemos pertenecido, y como tesista de la Maestría en Ciencias Sociales de la Escuela de Trabajo Social de la UNC. El libro puede ubicarse dentro de aquellos textos que se enfocan en las metodologías cuantitativas de investigación, y, particularmente, de aquellos que se centran en el método de encuesta y su principal técnica de producción, análisis e interpretación de datos: la estadística.

Uno de los objetivos principales de este manual es ofrecer una propuesta pedagógica y didáctica de utilidad para aquellos profesionales, estudiantes, docentes e investigadores, que necesiten realizar estudios cuantitativos en el área de las Ciencias Sociales y la Comunicación Social. Para ello, recurrimos, en gran parte, a lo planteado por diversos metodólogos, estadísticos, sociólogos, administradores de empresas y comunicadores sociales.

Consideramos que este trabajo es de carácter universitario y extensionista. Nos interesa particularmente democratizar y generalizar tanto las formas posibles de armar cuestionarios, aplicarlos y trabajarlos como las maneras estadísticas de procesar los datos provenientes de ellos. La generalización del uso de estas

herramientas podría coadyuvar a desarticular las pre-nociones existentes en el imaginario social respecto de estas estrategias de investigación.

Cual curso básico de Encuesta y Estadística, el objetivo fundante ha sido el de, dar a conocer el proceso que va desde el momento en que se decide realizar una encuesta hasta que se analizan los datos recabados para construir un informe final. De este objetivo inicial se desprenden varios específicos, relacionados con distintas etapas del proceso. En primer lugar, dar a conocer la importancia de la definición conceptual y operacional de las variables que se quieren medir. En segundo lugar, la manera de seleccionar los sujetos que se quieren estudiar. En tercero, la de construir y aplicar un cuestionario. En cuarto, el modo en que se trabajan, analizan e interpretan los datos obtenidos de las encuestas, a través del uso de algunas de las herramientas básicas que brinda la estadística. Y en quinto lugar, dar a conocer la manera en que se escribe un informe estadístico.

El texto consta de tres capítulos diferenciados aunque indisolublemente interrelacionados. En el primero, se desarrollan con detenimiento los conceptos de estadística, unidades de análisis, variable, valor, dato, tipos de variable, niveles de medición de las variables, muestreo (probabilística y no probabilístico). En el segundo, se detallan los pasos a seguir para construir un cuestionario. En este apartado se hace hincapié en las etapas de construcción y aplicación del instrumento de medición, las maneras de redactar las preguntas, los tipos de preguntas posibles, la etapa de codificación de los datos, la codificación física (o armado de la matriz de datos en un programa estadístico), la codificación manual. Luego, en el tercer capítulo se despliegan los conceptos y procedimientos básicos de la estadística descriptiva. Los contenidos de este apartado son: tablas de frecuencia, tablas de frecuencia para variables categóricas; tablas de frecuencia para variables cuantitativas; tablas de contingencia, medidas de tendencia central; utilidad de las medidas de tendencia central; medidas de dispersión; utilidad de las medidas de dispersión; algunas estrategias de redacción de informes estadísticos.

Antes de dar inicio al despliegue de los contenidos temáticos, queremos expresar nuestra posición en relación con los dos ejes que estructuran el texto: la encuesta y la estadística como herramientas de investigación en Ciencias Sociales y Comunicación Social.

La encuesta resulta en la aplicación de un procedimiento estandarizado de recolección de información a una muestra o población. El objetivo es indagar acerca de características socioeconómicas, demográficas, estructurales de los sujetos-objeto de estudio. Se trata de un método estructurado que requiere, en consecuencia, de una igual aplicación a todas las personas estudiadas (Sautu, Boniolo, Dalle y Elbert, 2005).

Una de las críticas que recaen sobre esta técnica es que la misma brinda información restringida sobre el objeto de estudio puesto que el *respondente* debe atenerse a una serie de respuestas predeterminadas. Del mismo modo, y en relación con esta crítica, suele argüirse que esta estrategia no contempla las referencias contextuales respecto de los individuos estudiados, hecho que limita la interpretación de los datos obtenidos (Sautu, Boniolo, Dalle y Elbert, 2005).

Si bien reconocemos estas limitaciones, consideramos aquí que la encuesta es una estrategia más en la investigación social y que la misma puede combinarse con naturalidad con otra serie de estrategias tanto cuantitativas como cualitativas a partir de la denominada “triangulación metodológica” (Vasilachis, 1992; Bourdieu, Chamboredon y Passeron, 2004). Las combinaciones de métodos de abordaje de la realidad, o “triangulación metodológica”, responden a la necesidad de abordar diferentes aspectos de la realidad social, los cuales obedecen a una naturaleza de carácter diferencial.

Si bien en Ciencias Sociales existe una multiplicidad de paradigmas de investigación, a modo ilustrativo hablaremos aquí de dos bien diferenciados: por una parte, el positivista; por la otra el interpretativista. Según el paradigma positivista (Sautu,

Boniolo, Dalle y Elbert, 2005) la naturaleza de la realidad es objetiva; la relación entre el investigador y el sujeto investigado es de distancia (no hay influencia de uno en otro); el investigador puede distanciarse de sus propias creencias, valores, ideologías durante el proceso de investigación; la metodología utilizada es básicamente cuantitativa, con un razonamiento deductivo para la elaboración del diseño de investigación y un razonamiento inductivo para el análisis de los datos recabados. Según este paradigma, es importante privilegiar las estructuras sociales que condicionan el accionar de los sujetos antes que la manera en que los sujetos interpretan su propia existencia. El representante más importante en Ciencias Sociales de este paradigma es Durkheim, considerado el “primer científico social” y el “primer sociólogo positivista” del siglo XIX (Corbetta, 2007, 11). Durkheim entendía que los hechos sociales debían ser entendidos como “cosas” (12). La realidad social debía ser entendida como externa al individuo y podía ser aprehendida objetivamente utilizando los métodos de las ciencias naturales. Para obtener conocimiento objetivo de la realidad, el sociólogo debía proceder de manera inductiva, pasando de la observación directa de hechos a la detección de regularidades que permitiesen establecer luego leyes universales (Corbetta, 2007). “La misión del científico” era entonces, la de reconocer “los principios organizadores de la realidad social” (12). La forma que debían adoptar las proposiciones o afirmaciones generales respecto de la realidad era la de las leyes causales. La “formalización matemática” (14) y el “experimento” (14) resultan en las metodologías predilectas de este paradigma. Las investigaciones centradas en encuestas, mayoritariamente, aunque no exclusivamente, responden a este paradigma de investigación.

Por otra parte, tenemos al paradigma interpretativo (Sautu, Boniolo, Dalle y Elbert, 2005), que entiende que la realidad es múltiple y subjetiva; que la relación entre el sujeto que investiga y el sujeto investigado es de mutua influencia e interdependencia; que los valores del investigador sí se ponen en juego al momento de investigar; que para abordar el objeto de estudio se trabaja más con metodologías cualitativas. El fundador de este paradigma en

Sociología y Ciencias Sociales fue Max Weber. Para Weber, la realidad social debe ser “interpretada” antes que observada (Corbetta, 2007, 20). El objetivo de la ciencia social era para Weber comprender racionalmente las motivaciones del accionar de los agentes, “entender la magnitud del propósito y la intención de la acción humana” (Corbetta, 2007, 21). Se trata más bien de un enfoque individualista, que busca determinar conexiones causales de hechos concretos, los cuales resultan en condiciones de posibilidad para la configuración de las acciones que se estudian (Corbetta, 2007). “A las leyes causales positivistas” Weber opone “enunciados y conexiones” caracterizados por la “especificidad y la posibilidad” (23). Como el objetivo es comprender el significado que los sujetos atribuyen a su propia acción, este paradigma orienta a los investigadores a utilizar métodos y técnicas cualitativas que impliquen una interacción empática entre el sujeto que investiga y el sujeto investigado

Si bien no es posible triangular, en las investigaciones, paradigmas epistémicos, sí lo es integrar armónicamente los enfoques cualitativos y cuantitativos a partir de la triangulación intramétodo o intermétodo; la triangulación de investigadores; la de fuentes; la de datos (cualitativos y cuantitativos); la triangulación de teorías y la multidisciplinaria (Gallart, 1993; Martínez Miguelez, 1999). En consecuencia, consideramos que el método de encuesta puede utilizarse en diferentes tipos de estudios en el área de las Ciencias Sociales y, en particular, en las Ciencias de la Comunicación, en combinación con otros métodos, sean éstos cualitativos o cuantitativos.

Asimismo sostenemos con Sautu, Boniolo, Dalle y Elbert (2005), Ghiglione (1989) y Hernández Sampieri, Fernández Collado y Baptista Lucio (2000) que la técnica de encuesta permite abarcar un amplio espectro de tópicos y contribuir a la generación de trabajos de carácter descriptivo de la realidad objeto de estudio. Entendemos aquí por estudios descriptivos a aquellos que “buscan especificar las propiedades importantes de las personas, grupos o comunidades o cualquier otro fenómeno que sea sometido a

análisis” (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000: 60). Cuando una realidad social, cultural, comunicacional ha sido explorada de manera intensiva, el investigador se encuentra habilitado para desarrollar un estudio extensivo. En función de las categorías de análisis construidas en una primera etapa exploratoria, es factible pasar a una segunda etapa, denominada “descriptiva”, que permita conocer cómo se manifiestan esas categorías o variables en una parte o en la totalidad de una población de interés. La encuesta es, en este sentido, uno de los métodos más apropiados para el logro de tal objetivo puesto que permite estudiar a muchas unidades de análisis en relativamente poco tiempo, y a un menor costo que si se utilizaran métodos cualitativos tales como las entrevistas en profundidad y/o antropológicas.

Por otra parte, y en consonancia con lo anterior, el instrumento de recolección de datos (o cuestionario) al ser estandarizado permite, además de lo anterior, una mayor celeridad en el tratamiento, análisis e interpretación de los datos recabados. Al mismo tiempo, la estandarización de las preguntas y opciones de respuesta facilitan la comparabilidad de los datos y permite evaluar la validez y fiabilidad del instrumento de medición (Sautu, Boniolo, Dalle y Elbert, 2005).

Como aventuramos en el párrafo anterior, la Estadística resulta en una herramienta importante para las Ciencias Sociales, no sólo porque permite generalizar resultados al total de una población, sino porque admite el estudio extensivo de pocas unidades de análisis.

La Estadística tuvo su ingreso en el campo de las Ciencias Sociales durante los comienzos de la filosofía positiva, cuando Comte (entre otros) traspoló el paradigma de conocimiento de las Ciencias Naturales al de las ciencias del hombre. Las Ciencias Sociales para adquirir el estatuto de ciencias debían liberarse de las especulaciones teológicas de la Edad Media y funcionar bajo los preceptos del método de las Ciencias Naturales. Los problemas

característicos de la segunda mitad del siglo XIX, propios del sistema capitalista, necesitaban de una solución, y de ello se encargarían las ciencias del hombre (tales como la sociología, la economía, el derecho, la psicología, la ciencia política, la lingüística), pero, para ello, el conocimiento científico de la realidad social debía ser racional, basado en la observación, medición, experimentación y comparación (De Luque, 1997).

Durkheim, el fundador del paradigma positivista en Investigación Social, a partir de sus obras maestras, *Las Reglas del Método Sociológico* y *El Suicidio* le imprimirá a la investigación un carácter positivista y cuantitativo.

El paradigma positivista bregó, en términos generales, por la realización de investigaciones empíricas y experimentos en los que se medían y se cuantificaban los resultados obtenidos. Algunos desarrollos teóricos precursores en las Ciencias de la Comunicación enmarcados dentro de este paradigma son, por una parte, la Teoría de la aguja hipodérmica, que entendía al receptor como directamente influenciado por parte de los medios; por la otra, los trabajos de Laswell en relación con la propaganda en el contexto del período entre guerras, que desarrollan la fórmula de las 5Q, y un marco teórico para la sociología funcionalista de los medios de comunicación (Mattelart, 1996; Wolf, 1996). Además, también pueden mencionarse los desarrollos de Lazarsfeld y Merton, con la Teoría de los efectos limitados y la noción de flujos comunicacionales a dos niveles ó two step flow (Wolf, 1996). Luego, los trabajos de Katz, que confluyeron en la conformación de la Teoría de Usos y Gratificaciones, a partir de considerar a la audiencia como parte activa del proceso de comunicación, y a los medios como algunas de las fuentes posibles de satisfacción de necesidades (Wolf, 1996; Mattelart y Mattelart, 1997). Ciertas líneas de investigación, ubicadas dentro de los estudios sobre los efectos de los medios son: la de indagación sobre los efectos sociales positivos y negativos del contenido de los medios; la de indagación sobre el modelado incubado de percepciones de la realidad social; la de la publicidad

como instrumento de socialización infantil, entre otros (Wimmer y Dominick, 1996). Al mismo tiempo, algunas otras líneas de investigación características de este paradigma, son: la indagación sobre los medios audiovisuales; la investigación sobre los medios impresos; la investigación sobre publicidad y relaciones públicas, entre otras (Wimmer y Dominick, 1996).

En la actualidad, se siguen realizando estudios en Comunicación Social que responden a estas u otras líneas de investigación que requieren de la Estadística como herramienta de tratamiento y análisis de la información.

En general, la Estadística suele ser entendida, desde el sentido común, como un conjunto incomprensible e indescifrable de números, porcentajes, cuadros y tablas (Grasso, 1999). La estadística se encuentra hoy desprestigiada por su incorrecta utilización por parte de distintos organismos de la sociedad, sean estos públicos o privados. Como corolario, el imaginario colectivo suele asociar la Estadística a un “artilugio” para “inventar” y “dibujar” datos (Grasso, 1999).

Por otra parte, la Estadística ha sido criticada fundamentalmente por su uso indiscriminado y acrítico por parte de muchos investigadores pertenecientes a la corriente positivista de investigación en Ciencias Sociales. Bourdieu, Chamboredon y Passeron (2004), en *El Oficio del Sociólogo*, dirán que el positivismo, que tiene una propensión a tomar a su objeto de estudio como si se tratara de una cosa, tiende a considerar como hechos los datos recabados sin tener en cuenta que cualquier dato que se obtenga de la realidad es el resultado de una construcción teórica previa respecto del objeto de estudio. Es decir, cualquier dato estadístico tiene significación sólo dentro de un sistema interrelacionado de conceptos previamente construido y basado en uno o más autores, los cuales conceptualizaron y definieron previamente esa realidad que se quiso estudiar. En síntesis, para los sociólogos, sin teoría no hay dato. Los datos son el resultado de una construcción teórica, reflexiva, y analítica. Además está decir que coincidimos con esta postura epistemológico-teórico-

metodológica de la sociología del conocimiento de Pierre Bourdieu.

Mientras tanto, y contrariamente a las nociones de sentido común, y a las realidades que han desprestigiado a la Estadística como herramienta de investigación, consideramos aquí que la misma es una técnica importante para la investigación social, y que no resulta ni en una mera colección de datos inútiles (Grasso, 1999), ni en una estrategia de manipulación de la opinión pública por parte del Estado, ni tampoco en un modo de conocer empirista contrapuesto al pensamiento teórico, reflexivo, analítico y crítico.

La estadística es una herramienta que permite a cualquier persona manejar una gran cantidad de datos, resumir la información con celeridad y detectar las tendencias que aquélla muestra. La/el investigador/a extrae porcentajes, promedios, desviaciones estándar y prepara el terreno para realizar una lectura e interpretación del dato. El significado final del dato estará dado por la definición teórica del objeto de estudio.

El presente libro tiene como intención desarrollar los procedimientos de descripción, resumen y análisis de datos pertenecientes a la Estadística descriptiva.

Para concluir consideramos fundamental expresar, siguiendo a Bourdieu, Chamboredon y Passeron (2004) que, en la investigación social, y muy particularmente en la investigación en Comunicación Social, la Estadística es una herramienta necesaria de objetivación y ruptura con las nociones de sentido común, que permite tomar conciencia de las estructuras sociales ajenas al conocimiento del individuo, las cuales funcionan como constreñimientos para su accionar. Esta herramienta luego se integra con otras, que responden a una lógica de trabajo de tipo cualitativa.

1. UNIVERSO, VARIABLES Y MUESTREO

1.1 Estadística y pensamiento estadístico

En términos de Blanch y Joeques (1997a) usualmente asociamos a la Estadística tanto con la recolección de grandes masas de datos y su presentación en tablas y gráficos, como con el cálculo de totales, promedios y porcentajes. Lo anterior es, según las autoras, sólo una parte de esta disciplina. La Estadística permite, además, diseñar experimentos y muestras, entre otra serie de acciones.

La Estadística es una herramienta de investigación que puede ser redefinida como un modo de pensar que permite resolver problemas sociales, problemas prácticos, de investigación científica y generar una ruptura con las ideas de sentido común que explican la realidad social. (Blanch y Joeques, 1997a; Bourdieu, Chamboredon y Passeron, 2004). En relación con el último punto, podemos decir que la Estadística como herramienta de investigación social es un instrumento de objetivación que permite desconcertar las primeras impresiones de la realidad objeto de estudio y construir nuevas relaciones entre hechos, fenómenos o variables. Desde un punto de vista sociológico y comunicacional, la Estadística permite, además, plantear hipótesis estructurales en torno al porqué de los fenómenos, no tanto en función de la manera en que los sujetos perciben la particular situación dentro de la cual están inmersos, sino de causas objetivas ajenas a su conciencia que constriñen su accionar.

Un análisis estadístico resulta en un conjunto de mediciones realizadas a determinado grupo de individuos u objetos. Los individuos u objetos a medir serán entendidos aquí como “unidades de análisis”. Las mediciones efectuadas a unidades de análisis (individuos) varían de una a otra (esto es, de un individuo a otro).

Supongamos que queremos realizar un estudio acerca del nivel de satisfacción del cliente en relación con el servicio brindado por una empresa X. El análisis estadístico: ¿sería necesario si, por ejemplo, todos los clientes tuvieran la misma percepción en relación a la calidad del servicio brindado por la empresa? La respuesta es no, puesto que el análisis estadístico de datos es de utilidad sólo cuando existe variabilidad entre las distintas mediciones realizadas (en el ejemplo planteado anteriormente, la variabilidad se expresaría en las diferentes apreciaciones de los distintos clientes de la empresa en relación con la calidad de servicio). Las unidades de análisis que estudiamos (es decir las personas), y por ende los datos (la información que recogemos respecto de las personas) son intrínsecamente variables (es decir, no hay dos personas exactamente iguales en todos los sentidos).

En síntesis, lo habitual es la variación entre las mediciones efectuadas. Y, siguiendo a Berenson, Levine y Krhebiel (2001), pensamiento estadístico es un conjunto de procesos de pensamiento (modos de pensar y razonar) que permiten comprender, administrar y reducir la variación de las mediciones efectuadas en las diferentes unidades de análisis.

1.2. Unidades de análisis, variables, valor y dato

Para comprender el concepto de Estadística y las modalidades de aplicación de esta forma de pensamiento, veremos ahora los conceptos de *variable*, *valor* y *dato*.

¿Qué es una variable? Es “cualquier característica que varía de una unidad de análisis a otra en una población objeto de estudio

o muestra” (Blanch y Joeques, 1997b, 8). Por ejemplo, los clientes o los públicos internos de una institución pueden ser clasificados en hombres y mujeres. La característica que estudiamos en este caso es la variable “sexo”. Esta variable está compuesta a su vez por dos aspectos o dimensiones: la dimensión “mujer” y la dimensión “varón”. Si realizáramos una encuesta, la característica sexo variará de un encuestado a otro. En una persona (unidad de análisis) la variable adopta la característica o valor *femenino* y en otra persona (unidad de análisis) la característica o valor *masculino*.

Podemos entonces ahora definir el concepto de *dato*: es el valor que adopta una variable en determinada unidad de análisis.

1.3 Tipos de variables

Para poder realizar las mediciones en las unidades de análisis previamente tenemos que conocer los tipos de variables que queremos medir.

Una clasificación diferencia entre variables categóricas o cualitativas y variables numéricas o cuantitativas.

- a) *Variables categóricas o cualitativas*: “son aquellas definidas por las categorías, clases o dimensiones que las componen” (Blanch y Joeques, 1997b, 8). Cuando hablamos de categorías, clases o dimensiones, hacemos referencias a características que pueden variar de una unidad de análisis a otra. Por ejemplo: Estado civil (categorías: casado, soltero, separado/divorciado; viudo, juntado), tipo de empleo/ocupación (categorías: empleado de comercio, ama de casa, profesión liberal; empleado público; etc.); nivel educativo alcanzado (categorías: primaria incompleta, primaria completa, secundaria incompleta,

secundaria completa, etc); nivel de conformidad con un servicio prestado (categorías: totalmente conforme, conforme, poco conforme, nada conforme). Como puede observarse, este tipo de variables tienen múltiples categorías que las definen.

Una sub-tipología de variables cualitativas o categóricas son las *variables dicotómicas o binarias*. Una *variable dicotómica o binaria* es aquella que se genera en función de la “presencia o ausencia de determinada categoría (o característica)” (Blanch y Joeques, 1997b, 8) en la unidad de análisis. Por ejemplo: si la variable es: “¿Compraría usted algún otro producto de esta empresa?”, y las categorías de respuesta son: “sí o no”. En la unidad de análisis hay presencia de la característica señalada (intención de comprar otro producto de la empresa) o ausencia de ella (ninguna intención de comprar otro producto).

b) *Variables numéricas o cuantitativas*: Estas variables “se expresan por un número” (Blanch y Joeques, 1997b, 9). Dentro de esta clasificación se enmarcan dos tipologías: las variables que se obtienen por simple conteo (variables discretas) y las que se obtienen a partir de mediciones (variables continuas).

Las *variables discretas* “surgen de un simple conteo o por asignación de ciertos códigos numéricos a las categorías de las variables cualitativas” (Blanch y Joeques, 1997b, 9). Por ejemplo: Número de años en el trabajo; Número de Hijos; Cantidad de revistas a las que está suscripto. Cantidad de clientes vip de una empresa. Cantidad de reclamos por el servicio

brindado. Importante es tener en cuenta que, cuando hablamos de variables discretas decimos que no hay valores intermedios entre un número y otro. Por ejemplo, no podemos decir que una familia tiene 1 hijo y medio: o tiene 1 hijo o tiene 2 hijos. No hay continuidad entre una unidad de análisis y otra. Son unidades separadas, discretas. Las contamos para saber cuántas hay.

Las *variables continuas*, en cambio, son aquellas que “surgen de un proceso de medición sobre las unidades de análisis” (Blanch y Joeques, 1997, 9). Ejemplo: Duración en minutos de llamadas telefónicas; Nivel de Ingresos; Edad; Altura, Peso, Temperatura. En este caso, lo importante es recordar que pueden existir infinidad de valores entre un valor y otro de la variable. Por ejemplo, podemos decir que una llamada duró tres minutos y veinte segundos. Entre el minuto tres y el minuto cuatro hay infinidad de valores intermedios. La llamada podría haber durado tres minutos y treinta segundos o tres minutos y cuarenta segundos, y así sucesivamente.

1.4 Grados de medición de las variables

Veamos ahora el *grado de medición* de las variables (Blanch y Joeques, 1997b).

Las variables anteriormente mencionadas pueden además, ser re-ordenadas según su grado de medición. Una clasificación para las escalas de medición de las variables es la siguiente: a) escala nominal; b) escala ordinal; c) escala intervalar; d) escala de razón. Vamos a las definiciones de cada tipo de escala:

La *escala nominal*, tal como su nombre lo indica, nombra, diferencia, identifica, clasifica a las unidades de análisis. Implica

la clasificación de las unidades de análisis según posean o no determinada característica cualitativa (Blanch y Joeques, 1997b). A este nivel de medición se asigna un valor arbitrario a cada categoría de la variable, aunque esto no implica establecer orden ni distancia numérica entre las distintas categorías. No se realizan supuestos acerca de las relaciones que pueden existir entre los valores de las variables. Veamos un ejemplo. La variable “estado civil del cliente” puede tener las categorías: soltero; casado; viudo; divorciado/separado; juntado. A cada categoría podemos agregarle un código numérico arbitrario, por ejemplo, a soltero el código 1; a casado el código 2; y así sucesivamente. Lo anterior, no significa que el cliente casado sea mejor o superior que el cliente soltero, ni que sea mejor cliente uno que el otro. Simplemente, se codifica para diferenciar una categoría de otra. Las unidades de análisis se agrupan, a este nivel, en clases exhaustivas y mutuamente excluyentes.

Todo lo anterior tiene una consecuencia inmediata para el tratamiento de los datos medidos en escala nominal. El análisis de los datos sólo puede emplear algunas operaciones matemáticas como el cálculo de las frecuencias absolutas, relativas de las categorías de las variables; cálculo de la moda y test de Chi Cuadrado.

La *escala ordinal* correspondiente con el siguiente nivel de medición, nombra y agrupa a las unidades de análisis en distintas categorías pero además establece una relación de orden entre los sujetos objeto de estudio (Blanch y Joeques, 1997b). “A este nivel se mide la intensidad, fuerza o tamaño con la que aparece una propiedad dada en cada uno de los objetos de investigación” (Maintz, Hölm y Hübner, 1992, 52). Se establece una graduación entre las unidades de análisis. La limitación de este nivel de medición radica en que no puede determinar con certeza cuánto mayor es una categoría respecto de otra, puesto que no mide distancias (Blanch y Joeques, 1997b). Por lo tanto, no podemos decir exactamente cuál es la distancia entre una unidad de análisis y otra. Por ejemplo, una variable ordinal es “clase social” (Blanch y Joeques, 1997b). Las categorías de esta variable pueden ser: baja, media y alta. Ahora

que tenemos las categorías de la variable, podemos codificarlas con números: 1 para “baja”; 2 para “media” y 3 para “alta”. Estas cifras designan el puesto en la gradación, son cifras de rango o números ordinales. Sabemos el puesto que ocupa cada unidad de análisis pero no la cantidad absoluta de una propiedad (clase social). Sabemos que un sujeto de clase media está por encima de otro de clase baja pero no sabemos exactamente cuánto por encima se encuentra. No podemos afirmar, por ejemplo, que el de clase media está 2 veces por encima de alguien de clase baja, los números son arbitrarios. Simplemente establecemos un orden entre las unidades de análisis.

En algunas ocasiones la escala de orden se construye en función de un acuerdo respecto a la distancia que separa una categoría de otra. Un ejemplo podría ser cuando un investigador quiere medir el grado de conformidad de la clientela de una empresa respecto de la calidad del servicio de atención al cliente. Aquí las categorías de respuesta de la variable bien podrían ser: “muy conforme”; “conforme”; “indiferente”; “disconforme”; “muy disconforme”. “Algunos investigadores probablemente considerarán esta escala de medición como ordinal y otros como intervalar”, dado el punto cero convencional (indiferente) (Blanch y Joeques, 1997b, 11).

Otros ejemplos de variables medibles a nivel ordinal son “máximo nivel de instrucción alcanzado” (primario incompleto, primario completo, secundario incompleto, secundario completo, universitario incompleto, universitario completo...); “Empleados” (obreros, supervisores, gerentes, directores) (Blanch y Joeques, 1997b, 11).

Los datos que surgen de variables medidas a un nivel ordinal deberán ser tratados de manera similar a aquellos correspondientes con un nivel nominal de medición.

La *escala de intervalo* “tiene la propiedad de asignar una medición de distancia entre los valores de la variable” (Blanch y Joeques, 1997b, 11). Cuando, por ejemplo, se plantea que una unidad de análisis está

por encima de otra, se puede especificar cuántas unidades de diferencia hay entre ambas. Se trata de un nivel de medición que incluye a los dos anteriores: el nominal y el ordinal. La escala se construye a partir de un cero artificial o convencional de origen, que representa la ausencia de la característica a medir. Estas disquisiciones, difíciles de comprender para los profesionales de las áreas de Ciencias Sociales y Humanas, surgieron en el seno de otras disciplinas. Otros ejemplos de variables medibles a nivel intervalar son “inteligencia, temperatura, tiempo, altura”. Blanch y Joeques (1997b) explican además que, por ejemplo, la escala centígrada de temperatura también es intervalar. Allí el cero es el punto de congelación del agua, y 100° el de ebullición. Si bien se trata de una escala con intervalos iguales, no se puede sostener que 18° es el doble que 9°.

Otro ejemplo de escala intervalar es la escala de Lickert con cero artificial.

Es una escala que se aplica a las variables cuantitativas continuas, de un nivel de medición de razón.

En este nivel de medición de variables, los estadísticos aplicables se amplían significativamente. De entre las medidas que se pueden extraer tenemos a “la media aritmética, la varianza, la desviación estándar.

La *escala de razón* “presenta todas las propiedades de orden y distancia de una escala de intervalo aunque adicionando un punto ‘0’ (o, lo que es igual, agregando un punto de origen)” (Blanch y Joeques, 1997b, 12). A este nivel de medición, las unidades son realmente equidistantes entre sí puesto que existe un punto cero natural.

Si bien en Ciencias Sociales no suelen utilizarse variables de este nivel de medición, encontramos en algunas disciplinas como la Economía, algunas tales como valores monetarios, población, tasas de recuperación. La variable “nivel de ingresos mensuales” responde a un nivel de razón.

Abunda este nivel de medición en las Ciencias Físico-Naturales. Por ejemplo, en Ciencias Químicas y/o Biológicas, en Ciencias Agronómicas o en Ciencias Físicas, podemos encontrar variables como las siguientes: cantidad de bacterias en una gota de leche; rendimiento de trigo por hectárea; distancia entre objetos (Blanch y Joeques, 1997b).

La presencia de un punto cero natural en la escala permite realizar comparaciones proporcionales. Según Baranger (1992), en este nivel de medición son posibles todas las operaciones matemáticas, incluidas las potencias, los logaritmos y las raíces cuadradas.

Para concluir esta sección, puede señalarse que conocer y distinguir los grados de medición de las variables, resulta imprescindible, puesto que el tipo de escala determina el tipo de tratamiento estadístico de los datos. Además, conocer el nivel de medición de las variables es fundamental para realizar encuestas y efectuar luego un adecuado tratamiento estadístico de datos.

1.5 Operacionalización de variables

1.5.1 Procedimiento de operacionalización de variables

Generalmente, cuando realizamos una encuesta tenemos inicialmente una idea acerca de lo que queremos conocer/estudiar/investigar. A esa idea tenemos que traducirla en una o varias frase/s o proposición/es que contenga/n una serie de variables que son aquello que vamos a medir en las unidades de análisis que seleccionemos para encuestar.

Supongamos que somos dueños de un periódico y queremos hacer una encuesta para medir el nivel de satisfacción del cliente en relación a la calidad del producto (diario, artículos publicados) y a la atención al cliente. En primer lugar, tenemos que tener en claro qué vamos a medir. Las variables principales en este caso son:

“nivel de satisfacción del cliente en relación con el producto” y “nivel de satisfacción del cliente en relación con el servicio de atención al cliente”. Otra variable que puede interesarnos sería conocer, por ejemplo, la modalidad de lectura del diario del cliente y su familia. Para ello tendríamos que averiguar la frecuencia con la que la persona lee el periódico; si su familia también lo lee (en caso de que la tenga); el tiempo que el sujeto dedica a leer el periódico.

Además, podemos también querer conocer el perfil social y económico de nuestra clientela. Para ello tendríamos que tener en cuenta las variables: estado civil; si tiene o no empleo; si es empleado full o part time; qué categoría o tipo de empleo tiene la persona; monto total de ingresos mensuales en la familia (variable que no siempre es posible de medida, por lo intrusivas que resultan preguntas de este tipo); máximo nivel educativo alcanzado, entre otros.

El proceso de construir cada variable con sus categorías respectivas para poder luego armar el cuestionario se denomina *Operacionalización de variables*. La operacionalización de variables es, en definitiva, “la reducción progresiva de un concepto relativamente abstracto a cierto número de conceptos-componentes menos abstractos y más limitados en su alcance, hasta hacer posible alcanzar los referentes en la realidad y especificarlos” (Greenwood, 1973, 87). Sin la operacionalización de las variables, es imposible elaborar los instrumentos de recolección de datos como el cuestionario.

1.5.1.1 Definición nominal del concepto a medir

La primera etapa en el proceso de operacionalización de variables consiste, según Korn (1968), Lazarsfeld (1969 y 1973), Maintz, Holm y Hübner (1992); Hernández Sampieri, Fernández Collado y Baptista Lucio (2000); Emanuelli (2009), en la definición real del concepto o variable a medir. El concepto, en el momento en que surge, “no es más que una entidad concebida en términos vagos” (Lazarsfeld, 1973, 36) y requiere, para ser convertido en preguntas de un cuestionario, de su especificación. Siguiendo con el ejemplo

citado anteriormente, podemos querer medir en nuestros clientes las variables “clase social” o “situación social y económica”; “nivel de satisfacción del cliente en relación con el producto”; “nivel de satisfacción del cliente en relación con el servicio”. Estos son los conceptos o variables que necesitan ser definidos. Tenemos que tener en claro qué queremos decir cuando hablamos de cada uno de estos conceptos. Por ejemplo, podemos definir a la situación social y económica de una persona como el lugar que ocupa en la sociedad en función de su máximo nivel educativo alcanzado, su tipo de ocupación actual, tipo de vivienda en la que vive.

1.5.1.2 Determinación de las dimensiones

La segunda etapa consiste en la definición operacional de los aspectos o dimensiones que componen el concepto general anterior. Se trata del análisis de los componentes de la primera noción (Lazarsfeld, 1973). Así, en el caso de la variable “posición social y económica” de una persona encontramos tres dimensiones o aspectos (que ya se encontraban en la definición nominal del concepto): “máximo nivel educativo alcanzado”; “tipo de ocupación actual”; “tipo de vivienda en la que vive” o “nivel de ingresos mensuales”.

1.5.1.3 Selección de indicadores

La tercera fase, consiste en la selección de indicadores de las dimensiones de los conceptos o variables en estudio.

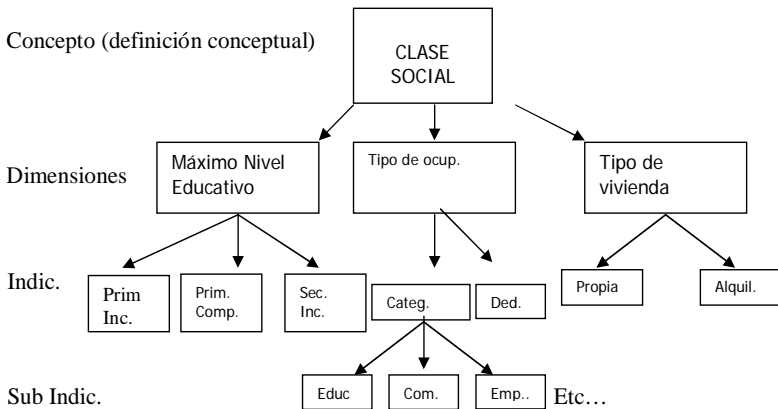
“Un indicador es una característica observable que puede existir o no en la unidad de análisis. Los indicadores empíricos provienen directamente de la experiencia de la vida cotidiana. Aquí la relación entre cada indicador y el concepto fundamental queda definida en términos de probabilidad y no de certeza, por lo que resulta necesario utilizar un gran número de indicadores” (Lazarsfeld, 1973, 36).

Veamos algunos indicadores para algunas de las dimensiones de la variable “posición económica y social de una persona”. De

la dimensión “máximo nivel educativo alcanzado”, tenemos los indicadores: primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, terciaria incompleta, terciaria completa; universitaria incompleta; universitaria completa. De la dimensión “tipo de ocupación actual”, podemos encontrar dos subdimensiones que a su vez pueden tener cada una de ellos sus propios indicadores. Una puede ser, por ejemplo, “¿trabaja?”, con dos indicadores posibles: 1. Sí; 2. No. Otras subdimensiones de la variable “tipo de ocupación actual” puede ser, “tipo de dedicación”, y algunos indicadores: 1. Part time; 2. Full time. Otra subdimension puede ser “categoría de ocupación”, con sus indicadores: 1. Educador; 2. Ama de casa; 3. Comerciante; 4. Empresario; 5. Ejecutivo; 6. Funcionario; 7. Profesión liberal; etc.

1.5.2 Ejemplo de operacionalización de la variable “clase social”

(Ejemplo tomado de Emanuelli (2002, 83) y modificado en función de una construcción propia del concepto seleccionado)



1.5.3 La operacionalización de variables en el contexto de una investigación científica. Su implicancia en el proceso de investigación y el armado de cuestionarios. Un ejemplo desde la Teoría de la acción social de Pierre Bourdieu.

Para armar un cuestionario resulta indispensable, si se trabaja en el marco de una investigación científica, haber construido previamente un “modelo de análisis” (Quivy y Compenhoudt, 1989) o “conjunto de hipótesis” el cual dé respuesta teórica a la pregunta planteada inicialmente en el diseño de investigación. Las hipótesis, o proposiciones que contienen variables, son construcciones teóricas derivadas del marco teórico conceptual del trabajo de investigación. Decimos entonces que el marco se conforma a partir de la teoría, esto es, a partir de un “conjunto interrelacionado de conceptos” (Sautu, Boniolo, Dalle y Elbert, 2005, 34) que permiten definir y comprender la realidad social, comunicacional, cultural, organizacional.

En las hipótesis de la investigación, se incluyen conceptos, muchos, todos o varios de los cuales ya están enunciados en el marco teórico conceptual. Estos conceptos pueden ser llamados, simultáneamente, *variables*. Las variables, deben ser operacionalizadas, es decir reducidas en su nivel de abstracción, para llegar a obtener los *indicadores* o elementos directamente observables en la realidad, que permitan, a su vez, construir el instrumento de recolección de información (por ejemplo, un cuestionario).

Veamos un ejemplo parcial de operacionalización de conceptos provenientes del marco teórico conceptual de una pequeña investigación en el área de las Ciencias Sociales. Si en un trabajo de tesis se plantea, como objetivo, estudiar la manera en que se desarrollan las trayectorias laborales de graduados de la carrera de Comunicación Social de una universidad, y se decide abordar las trayectorias desde el marco teórico conceptual desarrollado por Pierre Bourdieu, entonces uno de los desafíos será establecer una hipótesis.

La hipótesis surgirá de los desarrollos teóricos de Pierre Bourdieu que permitan dar una respuesta teórica a la pregunta de investigación. Para Bourdieu, las trayectorias de los agentes tienen relación con la evolución en el tiempo del capital social, económico y cultural de origen de los agentes estudiados (Bourdieu, 2006).

En la misma hipótesis, que es al mismo tiempo una definición conceptual, existen una serie de aspectos o dimensiones que habrá que reconocer y definir para encontrar los elementos empíricos que permitan determinar si esos aspectos del concepto *trayectoria* se manifiestan o no en la realidad y de qué manera. Las dimensiones encontradas en la hipótesis planteada anteriormente son: capital cultural; capital económico; capital social, evolución.

Veamos una posible operacionalización de una de las dimensiones: la de “capital cultural”. Pierre Bourdieu entiende por *capital cultural* a todo aquello ligado a conocimientos, ciencia y arte (Gutiérrez, 2005). El mismo puede existir, según Gutiérrez (pp.36-37), Bourdieu (2007, pp. 196-202) y Meischner (2007, pp. 4-5), bajo tres formas fundamentales: “en estado incorporado (bajo la forma de disposiciones durables relacionadas con determinado tipo de conocimientos, ideas, valores, habilidades); en estado objetivado (bajo la forma de bienes culturales, cuadros, libros, diccionarios, instrumentos); y en estado institucionalizado (como forma de objetivación como son los títulos escolares)”.

¿Cómo observo en la realidad concreta qué tipo de capital cultural mantiene una persona a la que quiero estudiar? Primero deberemos reconocer que la dimensión “capital cultural” del concepto “trayectoria laboral” está compuesta de tres subdimensiones: la sub-dimensión “estado incorporado”; la subdimensión “estado objetivado” y la subdimensión “estado institucionalizado”. Para cada subdimensión deberemos encontrar los indicadores o elementos directamente observables.

Algunos indicadores para determinar por ejemplo el “capital cultural en estado institucionalizado” de una persona pueden ser: “máximo nivel educativo alcanzado”: si primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, universitaria incompleta, universitaria completa; o los “títulos que ha obtenido”: por ejemplo, título de licenciatura; título de licenciatura y máster; título de licenciatura, máster y doctorado; título de licenciatura y doctorado; título de magisterio; título terciario, entre otros.

Algunos indicadores en relación con la subdimensión “capital cultural objetivado” pueden ser: posesión o no posesión de biblioteca; posesión/no posesión de cuadros. Algunos indicadores para la subdimensión “capital cultural incorporado” podrían ser el nivel de conocimiento respecto de algo (algunos indicadores posibles: respecto de la política, la historia universal, la historia argentina, la filosofía, o en general); ideología del agente; tipos de lectura habitual que la persona realiza (si lee revistas de divulgación científica; revistas de entretenimiento; revistas de arte, etc.); tipo de música que escucha semanalmente (si música folklórica, cuarteto; clásica, entre otras); actividades culturales a las que asiste cotidianamente.

Los indicadores son elementos importantes que permiten luego construir un instrumento de medición como el cuestionario para una encuesta. Estos indicadores responden a las dimensiones o subdimensiones de las dimensiones de una variable. Cada subdimensión puede ser convertida en una pregunta para un cuestionario, y los indicadores de cada subdimensión pueden ser convertidos de manera directa en opciones de respuesta a las preguntas (que responden a las subdimensiones).

Por ejemplo, una sub-sección de un cuestionario en el marco de la investigación sobre trayectorias de graduados de Comunicación Social, podría dedicarse a estudiar el nivel cultural de los agentes que se investigan. Para ello, se plantean preguntas (que hacen alusión a las subdimensiones del concepto “capital cultural”) y debajo de cada pregunta se incluyen las opciones de respuesta que son, en definitiva,

los indicadores que representan en los hechos a las subdimensiones de la dimensión “capital cultural” de la variable o concepto “trayectoria laboral”.

Indudablemente, aquí sólo esbozamos un ejemplo, y, en carácter de tal, no es exhaustivo y está sujeto a inexactitudes. Importante es destacar que cualquier pregunta en un cuestionario debe responder a una operacionalización de variables, que pertenecen a una o varias hipótesis, la/s cual/es son respuestas teóricas a preguntas de investigación en función de marcos teórico-conceptuales sumamente elaborados.

Siguiendo el ejemplo, y a modo aclaratorio, el procedimiento que se describe aquí para operacionalizar la variable “capital cultural”, deberá llevarse a cabo con cada una de las variables que conforman la respuesta teórica a la pregunta de investigación, es decir, el mismo procedimiento deberá aplicarse a las variables “capital económico”, “capital social” y “evolución”. Esto es lo que permite determinar cuáles son los datos que el investigador necesita recabar para su investigación (Quivy y Compenhoudt, 1989), y al mismo tiempo, construir instrumentos válidos de medición que realmente *midan* lo que inicialmente se pretendía (Maintz, Holm y Hübner (1992). De lo contrario, se corre siempre el riesgo de construir instrumentos de medición (u observación) que no respondan a los objetivos planteados en una investigación.

Las variables que se operacionalizan, probablemente puedan ser estudiadas a partir del diseño de instrumentos que respondan a las metodologías cualitativas (como temarios de entrevistas, por ejemplo). Otras variables, podrán ser estudiadas a partir de la técnica de encuesta. Lo fundamental es, en el segundo caso, construir la mayor cantidad de indicadores posibles por cada dimensión de cada variable para, de este modo, lograr armar buenos cuestionarios.

Un buen cuestionario está seccionado por áreas temáticas. Todo lo cual no significa que traslademos conceptos teóricos al cuestionario. Las preguntas en los cuestionarios deben ser redactadas en un lenguaje acorde a los sujetos-objeto de estudio, y deben evitar, en todo momento, trasladar palabras técnicas, conceptos y nociones que el investigador maneje en función de sus propias lecturas teóricas. Además, el encuestado no debe ser incomodado en ninguna circunstancia por una manera imprudente de formular preguntas.

En investigaciones académico-científicas, generalmente, una vez definidas las variables de las hipótesis se procede a la definición de la población total de individuos a estudiar, a la selección del método de muestreo, a la construcción y aplicación del cuestionario y, finalmente, al procesamiento, análisis e interpretación de los datos recabados.

1.5.4 Ejemplo de operacionalización de variables a partir de la Teoría de usos y gratificaciones. Implicancias para la construcción de un cuestionario.

La *teoría de usos y gratificaciones* se propone examinar la manera en que los consumidores hacen uso de los medios de comunicación masivos, al mismo tiempo que las satisfacciones que estos usuarios esperan obtener de su actividad consumidora de contenidos mediáticos (Wimmer y Dominick, 1996).

Los estudiosos que se enmarcan dentro de esta línea de investigación del campo disciplinar de la Comunicación Social entienden que la audiencia de los medios masivos es consciente de sus preferencias y capaz de hacer explícitas las razones o motivos que generan determinadas pautas de recepción de contenidos mediáticos. Es una línea que se enmarca dentro del paradigma positivista (Sautu, Boniolo, Dalle, Elbert, 2005), el cual entiende que la realidad es objetiva; que la relación que se establece entre el sujeto que investiga y el sujeto investigado es distante; que los valores del investigador

no interfieren en el proceso de investigación; y que las metodologías a utilizar son, mayoritariamente, cuantitativas. Es decir, se trata de una línea que permite desarrollar con mayor frecuencia estudios de tipo cuantitativos, con la aplicación de la técnica de encuesta.

En particular, y dentro de esta línea, podemos, por ejemplo, definir al concepto de “gratificaciones” como “lo que el consumidor espera satisfacer con su exposición a los medios de comunicación”. Es decir, definimos a las gratificaciones como si se trataran de “necesidades a satisfacer” por parte de los consumidores a partir del uso que éstos hacen de los medios de comunicación. Si hablamos de necesidades a satisfacer con el uso de las tecnologías de la comunicación e información, debemos todavía realizar un esfuerzo más para reducir el nivel de abstracción de la variable y encontrar sus dimensiones o aspectos, que no son más que los diferentes tipos de necesidades que un agente puede mantener.

Supongamos que queremos estudiar las gratificaciones que obtiene un grupo de personas de determinada institución a partir del uso de la pista de la información: Internet.

Podríamos definir *usos* como “la modalidad y frecuencia de acceso a la pista de la información”. Para el concepto “usos”, podemos delimitar los siguientes aspectos: tipos de páginas web más frecuentadas; frecuencia de acceso diario a Internet; frecuencia de acceso semanal a Internet; etc.;

Podríamos definir a *gratificaciones* como “las necesidades que la persona satisface a partir del uso de Internet”. Para el concepto gratificaciones, podríamos plantear las siguientes dimensiones: necesidades cognitivas; necesidades afectivas; necesidades recreativas; etc.

Ahora que tenemos las dimensiones de cada uno de los conceptos, procedemos a operacionalizar cada dimensión para encontrar los indicadores (o elementos directamente observables)

que permitan determinar si aquello que queremos estudiar se da o no en la realidad y de qué modo.

Entonces, buscamos primero los indicadores de la dimensión “necesidades afectivas”. Las necesidades afectivas pueden ser de relación con los amigos; de relación con la familia, entre otros.

Algunos indicadores para la dimensión más general “necesidades afectivas”, pueden ser:

- Uso Internet para contactarme con mis amigos
- Uso Internet para no sentirme sola/o
- Uso Internet para sentirme acompañada/o
- Uso Internet para contactarme con mis padres (a los que no veo seguido)
- Uso Internet para chatear con mi pareja; etc

Los indicadores, que pueden ser muchos, sirven para luego armar las categorías de respuesta en un cuestionario. La dimensión, que puede estar definida de manera técnica, teórica o conceptual, deberá ser traducida a una pregunta que pueda incluirse en el cuestionario antes de las categorías de respuesta. Por ejemplo, difícilmente en una encuesta preguntemos a los sujetos que estudiamos: “¿Cuáles son tus necesidades afectivas?” En todo caso preguntaremos: “En tu vida cotidiana y familiar ¿para qué utilizas Internet?” Luego, se pueden colocar las categorías de respuesta, que son los indicadores que surgen de la operacionalización de las variables. Importante es recordar que debemos evitar trasladar conceptos teóricos a las preguntas y/o indicadores (o categorías de respuesta) de los cuestionarios.

Otra dimensión que puede reducirse en su nivel de abstracción es: “Necesidades cognitivas”. En relación con las necesidades cognitivas podemos por ejemplo, encontrar los siguientes indicadores:

- Uso Internet para buscar información para mi trabajo
- Uso Internet para tomar conocimiento respecto de lo que sucede en mi ciudad

- Uso Internet para tomar conocimiento respecto de lo que sucede en mi país
- Uso Internet para tomar conocimiento respecto de lo que sucede en el mundo
- Uso Internet para estudiar y aprender nuevos conocimientos

La cantidad de indicadores es regulable, en función de las necesidades de nuestra investigación. Importante es aclarar que, cuando coloquemos, por ejemplo, las categorías de respuesta en un cuestionario, éstas no necesariamente son mutuamente excluyentes (en algunos casos, el *respondente* puede elegir sólo una opción de respuesta; en otros, puede elegir múltiples respuestas).

Una buena pregunta que anteceda los indicadores arriba descriptos y que pueda ser colocada en el cuestionario, podría ser: “¿Para qué utilizas Internet en tu vida laboral y académica?” Contrariamente, una pregunta errada sería: “¿Cuáles son tus necesidades cognitivas?”.

En relación con la variable “usos”, creemos que es más simple de operacionalizarla a partir de las dimensiones: “frecuencia de uso de Internet”; “tipos de páginas web más utilizadas”; “redes sociales más utilizadas”; etc. Dejamos al lector la libertad para que haga su propio ejercicio de operacionalización de tales dimensiones.

1.5.5 Operacionalización de variables en el contexto empresario/institucional

En estudios de mercadotecnia, marketing y en el marco de las empresas, generalmente los cuestionarios se construyen en función de determinadas necesidades concretas, y en muchas ocasiones, los mismos responden a modelos ya establecidos sobre cómo estudiar, por ejemplo, el clima laboral de una empresa o el perfil del consumidor, entre otros. Si bien no hay una armazón teórica que enmarque estos trabajos, sin embargo, existe la posibilidad de definir cada variable que se quiere estudiar, sea desde los manuales de calidad, los libros

de satisfacción al cliente, los artículos de investigación publicados en revistas nacionales o internacionales de prestigio.

Probablemente, los estudios realizados en el marco de una empresa puedan ser considerados *empiristas* desde la *sociología del conocimiento*. Ahora bien, imposible resulta, por ejemplo, exigirle a un *ejecutivo de cuentas* de una empresa a quien le encargan una encuesta en el plazo de tres días, que construya un marco teórico conceptual, puesto que semejante tarea puede durar entre seis meses y cinco años (o más) dependiendo del tipo de investigación de que se trate.

1.6 Población y muestra

Una vez que ya sabemos lo que queremos medir, tenemos que definir a quiénes vamos a estudiar, a qué sujetos vamos a encuestar. En el ejemplo de la empresa periodística, ésta podría querer conocer el nivel socioeconómico, el nivel de satisfacción en relación con el producto y el nivel de satisfacción en relación con la atención que mantienen todos sus clientes en un determinado momento. Al conjunto de todos los sujetos que conforman la clientela de la empresa podemos denominarlo, en este sentido, como *universo o población*. Población es, entonces, el conjunto de individuos en los cuales está interesado el investigador.

En ocasiones, cuando la población es muy grande, resulta imposible estudiarla en su totalidad. Por ello, se hace necesario establecer *muestras del total de la población*. Y ¿qué es una *muestra* de la población? Es una porción de sujetos que deberá ser representativa si se quiere generalizar los resultados al total de la población.

1.7 Muestreo

Para diseñar muestras debemos, en primer lugar, formular un *plan de muestreo* (Blanch y Joekes, 1997g). Éste establece los pasos a seguir para obtener una muestra luego de haber definido la población objetivo. A continuación, los pasos a seguir para concretar el plan de muestreo:

- 1) Definición de la población, de las unidades de muestreo, de la extensión del muestreo y del tiempo de recolección de la información.
- 2) Identificación del marco de muestreo.
- 3) Determinación del tamaño de la muestra.
- 4) Elección del método de selección de muestras y selección de la misma.
- 5) Definición del procedimiento de estimación y obtención de las estimaciones y errores de muestreo.

Veamos ahora en qué consiste cada uno de los pasos mencionados anteriormente.

1.7.1 Población, unidades de muestreo, extensión y tiempo

El paso 1, implica, en primer lugar, la identificación del marco de muestreo, es decir, la definición del conjunto de todos los sujetos que pueden salir seleccionados durante el proceso de selección de la muestra (Blanch y Joekes, 1997g). En el caso del ejemplo, la empresa editorial tendrá que describir a su población: todos los clientes registrados en la base de datos de clientes.

Por otra parte, este paso implica, además, definir cada una de las unidades de muestreo. Una unidad de muestreo es ó un elemento ó un conjunto de elementos de una población que puede/n salir seleccionado/s durante el proceso de muestreo (Blanch y Joekes, 1997g). En nuestro caso, dicha unidad es cada uno de los clientes de la empresa.

En tercer lugar, es fundamental establecer la extensión del muestreo, o lo que es igual, indicar el lugar donde se llevará a cabo. En el caso del ejemplo, y suponiendo que la empresa es de Buenos Aires, la extensión del muestreo será entonces, Capital Federal. Finalmente, también es necesario especificar cuándo se recabará la información, esto es, el tiempo. Podemos suponer que la empresa va a realizar la encuesta en el período de tiempo que va del 1º al 29 de octubre de 2011.

1.7.2 Marco de muestreo

Este paso implica identificar el marco de muestreo (Blanch y Joekes, 1997g). Es un listado de todas las unidades de observación (o sujetos) que pueden salir seleccionadas durante el proceso de selección de la muestra. En el caso del periódico, el listado podría obtenerse de la base de datos de clientes. El documento impreso con los nombres de todos los clientes de una empresa es un marco de muestreo adecuado. Dicho marco deberá asemejarse lo más posible a la población objetivo para evitar dejar fuera elementos de la población con determinadas características particulares. Por ejemplo, si una consultora de mercado se propone estudiar la intención de voto de los habitantes de una ciudad, deberá elegir un marco de muestreo que incluya a todos los sectores sociales. Si elige a una guía telefónica, dejará de lado a un gran sector poblacional que no posee teléfono aunque, probablemente, sí una intención de voto definida. Este último ejemplo ilustra el caso de elección inadecuada de un marco de muestreo.

1.7.3 Definición del tamaño muestral

El tercer paso requiere de la determinación del tamaño de la muestra, de los procedimientos de estimación, la obtención de las mismas y cálculo de su precisión (Blanch y Joekes, 1997g). Definir el tamaño muestral es decidir a cuántas personas (o unidades de análisis) vamos a encuestar. Extraer el tamaño de muestra no resulta en un cálculo tan sencillo. Puede extraerse el tamaño de una muestra

con el objetivo de estimar un parámetro de la población, a partir de un estadístico muestral o, por ejemplo, con el de estimar la porción de una población que tiene presencia o ausencia de determinada característica. A continuación desarrollamos tres modos de extraer el tamaño de muestra.

1.7.3.1 Estadístico muestral

Un estadístico muestral es un *estimador puntual*, es decir un “valor individual que se utiliza para aproximar un parámetro de población” (Triola, 2000, 289). Por ejemplo, la media de muestra (o promedio) es el mejor estimador puntual de la media de población μ .

1.7.3.2 Nivel de confianza/ intervalo de confianza

Para saber si el estadístico muestral, o lo que es lo mismo, el estimador puntual, es preciso o no, los estadísticos idearon los “intervalos de confianza”. Los intervalos de confianza consisten en “una gama de valores (o intervalo) que probablemente contienen al valor del verdadero parámetro poblacional” (Triola, 2000, 290) los cuales son estimados a partir de los datos surgidos de una muestra poblacional.

Entender lo anterior resulta fundamental para poder luego extraer el tamaño de una muestra.

Para definir dicho tamaño es necesario, en primer lugar, establecer el nivel de precisión deseado para la estimación. El nivel de precisión va a estar dado, particularmente, por el nivel de confianza y el error muestral.

Ahora bien, antes de hablar del nivel de confianza es necesario aclarar por qué fue necesario desarrollar esta metodología.

Hablar de nivel de confianza es hacer alusión directa a los intervalos de confianza. Los intervalos de confianza son necesarios para saber cuán bueno es el estimador puntual que hemos seleccionado. Si bien dijimos anteriormente que la media de una muestra es el mejor estimador puntual de una media de población, en realidad, esto queda supeditado al tamaño de la muestra. Si es muy pequeña, existe una probabilidad alta de que el valor de la media poblacional no necesariamente concuerde con el valor de la media muestral. Si tomáramos otra muestra de la misma población, probablemente obtendríamos un valor de media muestral diferente al valor de la media muestral de la primera muestra, y así sucesivamente. Sólo el promedio de todas las medias de muestras posibles sería igual a la media poblacional. Pero extraer todas las muestras posibles resulta imposible. Para resolver este inconveniente es que los estadísticos desarrollaron los *intervalos de confianza*.

Cada uno de los intervalos de confianza queda asociado además, a lo que los estadísticos denominan “grado de confianza”, esto es, a una medida de la certeza de que el intervalo contiene el parámetro poblacional (Triola, 2000, 290). El grado nivel o coeficiente de confianza se simboliza con $1-\alpha$ (alfa griega minúscula) y se expresa en el lenguaje de las probabilidades. Es una probabilidad expresada a menudo en términos porcentuales.

Los grados de confianza asociados a los intervalos que suelen utilizarse con mayor frecuencia son (Triola, 2000; Berenson y Levine, 1996; Blanch y Joekes, 1998):

- El 90% (con un α –o probabilidad de- = 0,10)
- El 95% (con un α = 0,05)
- El 99% (con un α = 0,01)

Para aquellos a quienes les resulte complejo entender de dónde provienen los porcentajes asociado al nivel de significación α van las siguientes aclaraciones:

- Para $\alpha = 0,10$, tenemos: $1 - \alpha = 1 - 0,10 = 0,90$
- Para $\alpha = 0,05$, tenemos: $1 - \alpha = 1 - 0,05 = 0,95$
- Para $\alpha = 0,01$, tenemos: $1 - \alpha = 1 - 0,01 = 0,99$

Además, el valor α puede dividirse en dos:

- $\alpha = 0,05$, $\alpha/2 = 0,05/2 = 0,025$
- $\alpha = 0,10$, $\alpha/2 = 0,10/2 = 0,05$
- $\alpha = 0,01$, $\alpha/2 = 0,01/2 = 0,005$

Lo anterior permite entender que la probabilidad de que los intervalos no contengan el parámetro poblacional es muy baja: de 0,05; ó de 0,10 ó de 0,01. Por el contrario, existe una alta probabilidad de que los intervalos sí contengan al parámetro poblacional: de un 0,90; ó de un 0,95; ó de un 0,99.

Gran parte de los trabajos de investigación utilizan un grado de confianza del 95% debido a que ofrece “un buen balance entre la precisión (reflejada en el ancho del intervalo de confianza) y la confiabilidad (expresada por el grado de confianza)” (Triola, 2000, 290).

Por último, una vez creado un intervalo de confianza, diremos por ejemplo:

“Quisimos establecer un intervalo de confianza con un 95% de confianza de que contenga al verdadero parámetro poblacional en relación con la edad promedio de los estudiantes universitarios correspondientes con determinada ciudad. El intervalo de confianza quedó conformado de la siguiente manera:

- Intervalo de Confianza de 0,95 (ó 95%) de la media de población (μ) es: 18 años $< \mu <$ 25 años.”

Lo anterior querría significar que, si seleccionáramos muchas muestras diferentes de por ejemplo, 100 sujetos, del total de la

población universitaria de una ciudad, y “construyéramos un intervalo de confianza similar del 95% para cada muestra, a la larga el 95% de esos intervalos contendría el verdadero parámetro poblacional μ ” (Triola, 2000, 291). Lo anterior no es igual a decir que existe un 95% de probabilidad de que el valor poblacional caiga dentro del intervalo. Esta última afirmación sería errónea puesto que μ es un valor fijo y no una variable aleatoria (Triola, 2000). Lo que se mide es la probabilidad de que el intervalo contenga o no al parámetro poblacional y no la probabilidad de que μ caiga dentro del intervalo. En todo caso, podríamos decir que el intervalo de confianza tiene un 95% de probabilidad de contener al verdadero parámetro poblacional.

1.7.3.3 Valor z

A cada grado de confianza se asocia un valor z.

Seguidamente, algunos de los grados de confianza más utilizados y sus valores z respectivos:

- 90% de confianza, con $\alpha = 0,10$, queda asociado al valor z $\alpha/2 = 1.645$
- 95% de confianza, con $\alpha = 0,05$, queda asociado al valor z $\alpha/2 = 1.96$
- 99% de confianza, con $\alpha = 0,01$, queda asociado al valor z $\alpha/2 = 2.575$

1.7.3.4 Error muestral

Si seguimos pensando en el ejemplo anteriormente planteado, en relación con la estimación de la media de edad poblacional del estudiantado universitario de una ciudad, podemos señalar que entre la media obtenida en una muestra y la media de la población total seguramente habrá una diferencia. Esta diferencia es denominada por los estadísticos como “error”.

El margen de error, cada vez que se utilizan datos de una muestra para estimar la media poblacional μ se define, siguiendo a Triola (2000, 293) como:

La máxima diferencia probable (con una probabilidad $1 - \alpha$) entre la media de muestra observada y el verdadero valor de la media de población μ . El margen de error E también se denomina error máximo de estimación y puede obtenerse multiplicando el valor crítico y la desviación estándar de las medias de muestra.

La fórmula para extraer el margen de error es la siguiente:

$$e = z (\sigma_x) / \sqrt{n}$$

Los elementos que componen esta fórmula son:

- Z = Valor crítico asociado al intervalo de confianza elegido
- σ_x = Desviación estándar de la población total, o lo que es lo mismo, la dispersión respecto de la media poblacional que mantiene la totalidad de datos (de la población total).
- \sqrt{n} = Raíz cuadrada de n, y n es el total de la muestra.

A continuación, la fórmula para extraer la desviación estándar de la población.

$$\sigma_x = \sqrt{\sum (x_i - \mu)^2 / N}$$

La fórmula quiere significar que la desviación estándar de la población total es igual a la raíz cuadrada ($\sqrt{\quad}$) de la sumatoria (Σ) de la resta de cada valor de la variable (x_i) menos la media poblacional μ elevada al cuadrado, y dividida (la sumatoria Σ) sobre N, es decir, sobre el total de la población.

Cuando se desconoce la desviación estándar de la población no hay que desesperar. Si bien el cálculo del margen de error “E” requiere, por parte del investigador, de conocer la desviación estándar de la población “ σ_x ”, en la mayoría de los casos esto no es posible porque tampoco se conoce μ (para calcular la desviación estándar de la población es necesario primero haber calculado la media poblacional) (Triola, 2000; Berenson y Levine, 1996).

En consecuencia, si desconocemos los parámetros poblacionales, podemos tomar una muestra mayor a 30 sujetos y extraer tanto la media como la desviación estándar muestrales para, a continuación, sustituir σ_x de la fórmula original por la s de la muestra.

Si el investigador, a partir de datos provenientes de estudios anteriores, conoce ya una media muestral y la consecuente desviación estándar, podrá utilizar estos datos para el cálculo del error (Triola, 2000).

1.7.3.5 Tamaño de muestra para la media

Para conocer el tamaño de muestra la fórmula es la siguiente:

$$n = (Z)^2 (\sigma_x)^2 / e^2$$

Es decir,

- $(Z)^2$ = Valor crítico elevado al cuadrado, el cual depende del nivel de confianza deseado.
- $(\sigma_x)^2$ = desviación estándar de la población elevada al cuadrado
- e^2 = error de muestreo permitido elevado al cuadrado

Para saber cuáles son los valores o cantidades correspondientes con cada uno de estos estadísticos de la fórmula, lo correcto sería consultar con una persona experta, es decir, con aquel/aquella que

conoce en profundidad el comportamiento de las variables que se pretenden estudiar (Berenson y Levine, 1996).

Sin embargo, y como esto no siempre es posible, a continuación se esbozan algunas sugerencias:

1. Si bien suele utilizarse más frecuentemente un intervalo de confianza de un 95%, en algunas ocasiones, puede ser necesario utilizar uno mayor: de un 99%. Mientras tanto, en otras ocasiones, puede resultar práctico reducir el nivel de confianza a un 90% (si no es indispensable tener tanta certeza) (Berenson y Levine, 1996).
2. En relación con el error de muestreo, se puede pensar en términos de cuánto error estamos dispuestos a tolerar para obtener, a pesar de él, resultados válidos (Berenson y Levine, 1996).
3. En relación con la desviación estándar de la población, no siempre la conocemos. Esto es, no en todos los casos conocemos cómo se distribuyen los valores que asume una variable alrededor de la media poblacional. Si fuera ése el caso, resulta conveniente realizar una estimación de la desviación estándar a partir de datos anteriores o un estudio piloto con una muestra de población para estimar una desviación estándar que permita realizar el cálculo del tamaño de muestra (Berenson y Levine, 1996).

1.7.3.6 Tamaño de muestra para poblaciones finitas

Una población finita es aquella que incluye un número limitado de observaciones y/o medidas (Berenson y Levine, 1996). Cuando trabajamos con poblaciones pequeñas y queremos extraer una muestra es necesario aplicar el “factor de corrección para poblaciones finitas”.

Si el lector recuerda, en el apartado 1.7.3.1 utilizamos la fórmula:

- $n = (Z)^2 (\sigma_x)^2 / e^2,$

Ahora denominaremos a n como n_0 . Llamaremos a n_0 como tamaño de muestra sin factor de corrección.

La fórmula ahora, con factor de corrección será:

$$n = (n_0) N / n_0 + (N - 1)$$

Lo anterior significa:

- n_0 = Tamaño de muestra sin factor de corrección
- N = Tamaño de la población (finita, se conoce)
- $(N - 1)$ = Tamaño de la población menos 1

1.7.3.7 Tamaño de muestra para una proporción

Supongamos que el director de una revista científica quiere conocer el tamaño de muestra necesario para estimar la proporción de la población de total de artículos publicados que tienen errores en los modos de citar la bibliografía.

El error de muestreo, e , será igual a la diferencia entre la porción de la muestra (p_s) y el parámetro que se va a estimar (p). La fórmula es la siguiente:

$$e = Z \sqrt{p(1-p) / n}$$

Z= Nivel crítico

p= porción verdadera de éxitos en la muestra considerada

n= total de muestra

Fórmula para extracción del tamaño de la muestra:

$$n = Z^2 p (1-p) / e^2$$

Los elementos que conforman la fórmula son:

Z = nivel de confianza deseado

e^2 = error de muestreo permitido

p= proporción de éxitos

Si sabemos el grado de confianza, podemos determinar el valor adecuado de Z de la distribución normal.

Recordando:

- 90% de confianza, con $\alpha = 0,10$, queda asociado al valor crítico $z_{\alpha/2} = 1.645$
- 95% de confianza, con $\alpha = 0,05$, queda asociado al valor crítico $z_{\alpha/2} = 1.96$
- 99% de confianza, con $\alpha = 0,01$, queda asociado al valor crítico $z_{\alpha/2} = 2.575$

Para estimar la probabilidad de éxitos se puede acudir a información pasada que nos brinde “una estimación estudiada de p” (Berenson y Levine, 1996, 366). Si no hay disponible información relevante, entonces proporcionamos un valor p de manera convencional. El valor que se obtiene de la multiplicación $p(1-p)$ se encuentra en el numerador de la fórmula, motivo por el cual este número debe ser lo más grande posible. Está comprobado que cuando $p=0,5$, el producto $p(1-p)$ logra su valor máximo. Entonces, cuando no contamos con “una estimación previa de la

porción verdadera, p , podemos utilizar $p=0,5$, como la forma más conservadora de determinar el tamaño de muestra” (367).

Algunos valores p junto con el correspondiente valor del producto $p(1-p)$ son (Berenson y Levine, 1996, 366):

- $p= 0,5$, $p(1-p)= (0,5)(0,5) = 0,25$
- $p= 0,4$, $p(1-p)= (0,4)(0,6) = 0,24$
- $p= 0,3$, $p(1-p)= (0,3)(0,7) = 0,21$
- $p= 0,1$, $p(1-p)= (0,1)(0,9) = 0,09$
- $p= 0,01$, $p(1-p)= (0,01)(0,99) = 0,0099$

1.7.4 Elección del método de selección de muestras

El paso número 4 consiste en decidir qué modalidad de muestreo se va a emplear. (Blanch y Joeques, 1997g). En este sentido, podemos distinguir dos tipos de métodos de selección de muestras. Uno de ellos se denomina “muestreo probabilístico” y el otro “muestreo no probabilístico”. El muestreo probabilístico, implica que cada uno de los sujetos que componen la población-objeto de estudio tiene una probabilidad conocida de salir seleccionado en la muestra. El muestreo probabilístico obedece a la lógica de la teoría de probabilidades. Una de las ventajas principales del muestreo probabilístico es que en función de los resultados obtenidos para la muestra podemos hacer inferencias o generalizaciones sobre el estado población objetivo.

En el muestreo no probabilístico, las unidades de análisis o sujetos de la población se eligen en función del criterio del propio investigador, motivo por el cual se desconoce la probabilidad que tiene cada sujeto de ser seleccionado para la muestra. En estos casos se desconoce la población total.

1.7.4.1 *Muestreos probabilísticos*

Dentro de los muestreos probabilísticos tenemos tres que son los que se aplican con mayor frecuencia: el muestreo aleatorio simple, el muestreo sistemático y el estratificado.

1.7.4.1.1 Muestreo aleatorio simple

Supongamos que una empresa editorial quiere estudiar la imagen que los públicos internos (empleados) mantienen de ella misma. Supongamos, además, que se trata de una gran empresa, con un total de 8000 empleados cuyos nombres y datos personales se encuentran en un fichero, y que, además, las fichas están ordenadas alfabéticamente y numeradas desde 1 hasta 8000. La empresa define a la población como a la totalidad de empleados de la empresa. Luego, reconoce que su marco de muestreo es el fichero, y que las unidades de muestreo son cada uno de los empleados. Define además que quiere encuestar a 100 empleados del total de la población objetivo. La selección de la muestra puede ser:

- a) A partir del uso de un software estadístico (EXCEL, INFOSTAT, entre otros)
- b) Por sorteo o tómbola de los elementos constitutivos de la población (Emanuelli, 2002): este método consiste, en primer lugar, en la enumeración de la totalidad de los elementos que componen la población. En segundo término, en anotar los números en papelitos (colocando esos papelitos en una caja) o colocar en un bolillero la totalidad de bolillas (de 1 hasta N). En tercer lugar, en extraer los elementos uno a uno hasta completar n (entendiendo a n por tamaño de muestra. En el ejemplo n =100). Si trabajamos con bolillero, y sale primero la bolilla 10, esto significa que el sujeto ubicado en el lugar 10 del fichero o base de datos queda seleccionado para la muestra. Si luego sale la bolilla número 5500, esto quiere decir que el sujeto que aparece en el lugar 5500 del fichero

también queda seleccionado. Y así se siguen seleccionando sujetos hasta completar la muestra.

- c) Por la utilización de tablas aleatorias: este método de selección requiere no sólo conocer el tamaño y el marco de la muestra, sino también disponer de una tabla de números aleatorios. Las tablas aleatorias son, según Emanuelli (2002, 91 y Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000, 217), “un compuesto de múltiples combinaciones de números extraídos por sorteo generados por una ruleta electrónica donde cada dígito tiene una probabilidad igual de aparecer en cualquier lugar de la lista”. Ahora bien, ¿cómo se utiliza la tabla de números aleatorios? Para el caso analizado, primero debemos determinar la cantidad de dígitos del tamaño de la población. En este caso, el número 8000 tiene 4 dígitos. En segundo lugar, si trabajamos con tablas aleatorias que, por ejemplo, presentan los números en columnas de a dos dígitos, deberemos reunir los dos números de la primera columna con los dos números de la segunda para construir una nueva que incorpore cuatro dígitos. Lo mismo deberemos hacer con las siguientes para contar con columnas de cuatro dígitos. En tercer lugar, se hará necesario numerar las nuevas columnas que se han conformado a partir del paso anterior. En cuarto lugar, y para iniciar la selección, hay que sortear un número entre las columnas del 1 al 20 (supongamos que sale sorteado el número 7) y otro número entre las filas del 1 al 10 (para el ejemplo supongamos que sale sorteada la fila 1). En la intersección de la columna y la fila que salieron sorteadas se ubica el número de 4 dígitos inicial, que representa al primer sujeto que conformará la muestra. Para no olvidarnos de este número inicial resulta conveniente subrayarlo en la tabla. En la tabla aleatoria que se presenta abajo, el número que se corresponde con la columna 7 – fila 1 es 3548. Desde allí partirá la selección de tantos elementos como el tamaño de n (en este caso, 100 sujetos). Se sigue hacia abajo y el

número siguiente, si no supera a N (a 8000) también queda seleccionado. En la tabla aleatoria de abajo, el segundo número, debajo del 3548 es 5240. Como no es superior a 8000 entonces se lo selecciona. Y así sucesivamente. Si al continuar aparecen números superiores a 8000, entonces aquellos se dejan de lado y se sigue con el número siguiente, y así sucesivamente hasta completar n . Recordamos que cada número que sale seleccionado en la tabla se corresponde con un sujeto que forma parte del marco de muestreo. Es decir, el número 3548 se corresponde con el sujeto 3548 del marco de muestreo y así sucesivamente. La selección seguirá por la columna 7 y cuando se llegue al final de la tabla se pasará a la columna 8 fila 1 para seguir la selección. Es decir, empezamos la selección en la columna 7 fila 1; seguimos por el dato que le sigue, ubicado en la intersección columna 7 fila 2; luego será el dato ubicado en la intersección columna 7 fila 3 y así sucesivamente hasta llegar a la última fila de la columna 7. De allí pasamos a la columna 8 fila 1, luego a la columna 8 fila 2 y así sucesivamente.

1.7.4.1.2 Muestreo sistemático

	Columna 1	Columna 2	Columna 3	Columna 4	Columna 5	Columna 6	Columna 7
F 1	1 0 0 9	7 3 2 5	3 3 7 6	5 2 0 1	3 5 8 6	3 4 6 7	3 5 4 8
F 2	3 7 5 4	2 0 4 8	0 5 6 4	8 9 4 7	4 2 9 6	2 4 8 0	5 2 4 0
F 3	0 8 4 2	2 6 8 9	5 3 1 9	6 4 5 0	9 3 0 3	2 3 2 0	9 0 2 5
F 4	9 9 0 1	9 0 2 5	2 9 0 9	3 7 6 7	0 7 1 5	3 8 3 1	1 3 1 1
F 5	1 2 8 0	7 9 9 9	7 0 8 0	1 5 7 3	6 1 4 7	6 4 0 3	2 3 6 6

Es una técnica muy utilizada por la sencillez del procedimiento y porque presenta ciertas ventajas adicionales (en relación con el muestreo aleatorio simple) (Emanuelli, 2002). Aquí, el investigador fija la posición relativa de las unidades incluidas en la muestra, lo que asegura una representación uniforme de los elementos de la población (Blanch y Joeques, 1997g). Como el método nos asegura

una muestra regularmente espaciada, hay mayor probabilidad de obtener una estimación más precisa de la medida de la población.

Supongamos que en una empresa la gerencia de Comunicación Institucional quiere averiguar cómo evalúan los empleados al clima laboral. El total de empleados es de 500 y se decide tomar una muestra sistemática de 150. La población en este caso está conformada por los 500 empleados de la firma. El marco de muestra puede ser el documento impreso (extraído de la base de datos de empleados) donde figuran los nombres de todos los trabajadores ordenados alfabéticamente.

En primer lugar, se enumera a la población de 1 hasta N (siendo N el total de la población, en este caso, 500). Luego se extrae una fracción de muestreo N/n , esto es, en el caso particular que analizamos, $500/100= 5$. A esta fracción de muestreo la vamos a denominar como $k=5$. La fracción de muestreo indica que tendremos que seleccionar un sujeto de cada 5 unidades de análisis.

A continuación, se sortea un número entre los 5 primeros correspondientes al primer intervalo. Así determinamos el punto de inicio y arranque r . Supongamos que sale sorteado el 5. Entonces vamos a la tabla y buscamos el número correspondiente con la fila 5. Ese número, que se va a corresponder con un sujeto de la muestra, queda seleccionado. Luego, deberemos saltar 5 números para elegir el siguiente (que representa también a un individuo de la muestra). Entonces, seleccionamos al sujeto que aparezca 5 posiciones por debajo del primero; luego, el próximo que se encuentre 5 lugares por debajo del segundo sujeto elegido, y así sucesivamente hasta completar el tamaño de la muestra (en este caso, de 100 sujetos).

Al igual que el muestreo aleatorio simple, este tipo de muestreo se puede realizar hoy de manera mucho más sencilla a través de softwares estadísticos tales como INFOSTAT, EXCEL, SPSS, STATA.

1.7.4.1.3 Muestreo estratificado

Para extraer un muestreo estratificado es necesario seguir los siguientes procedimientos (Blanch y Joekes, 1997g):

- 1) Determinar estratos o subgrupos de la población de interés.
- 2) Conocer la población total (dentro de la cual se encuentran los diferentes estratos de población).
- 3) Seleccionar dentro de cada estrato una muestra aleatoria simple.
- 4) Calcular el estadístico de interés (si hay interés en estimar parámetros poblacionales). Para estimar el parámetro de la población total habrá que ponderar las medidas obtenidas por cada estrato.
- 5) Calcular para cada estrato las varianzas (si se requiere de este cálculo). Para estimar la varianza poblacional habrá que ponderar las varianzas obtenidas para cada subgrupo de la población.

¿Cómo hacemos para establecer los estratos de población? Éstos “deben ser mutuamente excluyentes y conjuntamente exhaustivos” (Blanch y Joekes, 1997g, 55). Los estratos definen todas las categorías posibles de la variable de estratificación (es decir, determinan la exhaustividad) y establecen el hecho de que cada elemento de la población puede ubicarse sólo en uno de los estratos (es decir, cada estrato es mutuamente excluyente).

Por ejemplo, la población total podría dividirse en función del sexo de los individuos. En este sentido, tendríamos dos estratos, uno compuesto sólo por mujeres y otro por hombres. Ambos estratos definirían el dominio completo de la variable “sexo” (los estratos o

categorías son solamente dos), y al mismo tiempo, nadie asignaría una mujer a la categoría varones; ni un hombre a la categoría mujeres (cada elemento de la población puede ser asignado solamente a un estrato) (Blanch y Joeques, 1997g).

1.7.4.1.3.1 Utilidad del muestreo estratificado

El muestreo estratificado es sumamente útil cuando:

- a) Se trabaja con determinados subgrupos poblacionales que requieren, para su abordaje, de precisión. En estos casos, resulta conveniente considerar a cada estrato como si se tratara de una población (Blanch y Joeques, 1997g).
- b) Se requiere organizar el proceso de recolección de datos según la conveniencia del investigador. Por ejemplo, si se quiere estudiar la manera en que los habitantes de determinada ciudad resignifican los contenidos mediáticos televisivos, probablemente resulte conveniente tomar en consideración las condiciones materiales de existencia de los sujetos, dividiendo a la población en estratos según zonas o barrios (Blanch y Joeques, 1997g).
- c) Se presentan dificultades por la gran variación existente entre diversos sectores de una población. El procedimiento de muestreo puede variar según sean los estratos conformados. Por ejemplo, las personas que viven en instituciones como hospitales, asilos, serán más fáciles de encuestar (en su totalidad) que aquéllas que viven en casas particulares. En ocasiones, se suele encuestar a todas las personas que conforman un estrato y muestrear aquéllas que conforman otro (Blanch y Joeques, 1997g).
- d) Se requiere de estimaciones precisas respecto de las características de la población total. Para lograr tal objetivo

se tienden a conformar subpoblaciones más homogéneas a partir de una población heterogénea. A más homogéneo el estrato, menor variabilidad en las mediciones obtenidas en las unidades de análisis. Esto permite obtener estimaciones precisas a partir de una pequeña muestra de cada estrato, y la combinación ponderada de estimaciones puntuales permitirá realizar una estimación precisa de los parámetros de la población total (Blanch y Joeques, 1997g).

1.7.4.1.3.2 Definición de la población

Si quisiéramos estudiar, por ejemplo, desde la teoría de la Agenda Setting los temas sobre los cuales hablan los estudiantes de una facultad de comunicación de determinada universidad, tendríamos que conocer la totalidad de alumnos de esa facultad. Supongamos que hay un total de 50 estudiantes, que hay 9 en el primer año; 10 en segundo año; 8 en tercer año; 11 en cuarto y 12 en quinto.

1.7.4.1.3.3 Identificación del marco de muestra

Para el ejemplo anterior, el marco de muestra sería el registro de alumnos donde encontraríamos, además del nombre de cada uno, información complementaria referida a su trayectoria académica, condiciones sociales y/o económicas.

1.7.4.1.3.4 Selección de la muestra

En este muestreo, la población total N se divide en subpoblaciones de $N_1, N_2, \dots, N_h, \dots, N_k$ elementos respectivamente, de modo tal que:

$$N_1 + N_2 + N_h + \dots + N_k = N$$

Una vez especificados los estratos de población, se extrae una muestra de cada uno. La elección de la muestra para cada estrato se realiza de manera independiente (Blanch y Joeques, 1997g). Los

tamaños de muestra dentro de los estratos se simbolizan como: n_1 , n_2 , ..., n_h , ..., n_k respectivamente.

Por ejemplo, en el caso que venimos planteando, podemos tener interés en estudiar las diferencias entre los temas (o noticias) de los que habitualmente hablan los estudiantes de los primeros tres años y aquéllos de los que hablan los de los últimos dos años.

Al agrupar a los alumnos en función de si se encuentran en los primeros tres años o en los dos últimos, estamos creando los estratos.

Dijimos que había 9 estudiantes en primer año; 10 en segundo y 8 en tercero. La suma total es de 27 alumnos. Es decir, el primer estrato, N_1 tiene un total de 27 sujetos.

Por otra parte, dijimos que había 11 alumnos en cuarto año y 12 en quinto, es decir, en total, en estos dos últimos años, 23 alumnos. Para el segundo estrato, N_2 , hay 23 alumnos.

Supongamos que del estrato N_1 se decide extraer una muestra n_1 de 3 personas y del estrato N_2 se decide extraer otra muestra de n_2 de 2 personas.

Si por ejemplo, el total de sujetos que conforman el estrato N_1 fueran 30 y el total de sujetos que conforman el estrato N_2 fueran 20, y hubiésemos extraído una muestra n_1 de 3 sujetos y otra muestra n_2 de 2 sujetos, estaríamos ante un muestreo estratificado proporcionado, puesto que “los elementos de la muestra total (de 5 sujetos en este caso) son asignados a los estratos en proporción a la cantidad de elementos que poseen los estratos en la población” (Blanch y Joeques, 1997g, 58).

En notación estadística, lo anterior significa lo siguiente:

$$\frac{n_h}{n} = \frac{N_h}{N} \quad (\text{ponderación del estrato})$$

1.7.4.2 Muestréos no probabilísticos

Los muestréos no probabilísticos son aquellos en los que el investigador elige intencionalmente a las personas a las cuales va a estudiar. Algunos de los más usuales son el muestreo por conveniencia, el muestreo por juicio del investigador, el muestreo bola de nieve y el muestreo por cuotas.

1.7.4.2.1 Muestreo por conveniencia

En el muestreo por conveniencia el investigador selecciona a los sujetos-objeto de estudio según su propia conveniencia (Blanch y Joekes, 1997g). Por ejemplo, un comunicador social puede tener interés en conocer la opinión de las personas que se encuentran en una exposición de fotografías respecto de un determinado programa de televisión. O un comunicador institucional puede querer conocer la opinión de un grupo de personas que se concentraron en un determinado momento en un sector de la empresa respecto de cómo entienden al clima laboral. Este tipo de muestreo se suele hacer de manera espontánea y no permite generalizar los resultados de la muestra al total de la población (el investigador desconoce la composición de la población total).

1.7.4.2.2 Muestreo por juicio del investigador

El muestreo por juicio o “*purposive sampling*”, “*criterion based selection*” (Maxwell, 1996, 6); (Le Compte y Preissle, 1996, en Maxwell) implica la selección de sujetos u elementos de una población siguiendo un criterio específico, pautado por el propio investigador, en función de sus propios intereses teóricos y de investigación. Se trata de una estrategia que permite seleccionar de manera deliberada personas, eventos o escenarios particulares con el objetivo de obtener información relevante a los fines de la investigación.

El investigador selecciona según criterios teórico-metodológicos determinadas unidades de análisis puesto que son

éstos los que permitirán dar respuesta a su pregunta de investigación. Por ejemplo, un investigador puede querer conocer las diferentes trayectorias laborales de comunicadores sociales egresados en determinado año que se desempeñan hoy en ámbitos institucionales diversos, buscando conocer las similitudes y diferencias según sean mujeres o varones. Como trabaja desde el enfoque desarrollado por Pierre Bourdieu, delimita de manera teórica los ámbitos en los que podría localizar a diferentes sujetos para su estudio. Pierre Bourdieu habla de campos sociales, los define como “espacios de juego históricamente constituidos con sus instituciones específicas y sus leyes de funcionamiento propias” (Bourdieu, 1987, en Gutiérrez, 2005, p.31). Los mismos se presentan a la aprehensión sincrónica como “sistemas de posiciones y de relaciones entre posiciones” (Costa, 1976, en Gutiérrez, 2005, p.31), a los cuales están ligados cierto número de propiedades analizables independientemente de las características de quienes las ocupan. Estas propiedades son: los intereses propios de ese campo (o el capital específico que estructura el campo); una distribución dada del capital específico que se disputa (el cual ha sido acumulado en el curso de luchas anteriores y orienta en la actualidad las estrategias de los agentes); agentes que reconocen las reglas del juego y lo consideran digno de ser jugado; una lucha por conservar o transformar el campo de fuerzas establecido. Así Bourdieu entiende que existen diferentes campos sociales tales como el campo académico, el campo científico, el campo mediático, el campo escolar, el campo estatal. Al mismo tiempo, algunos estudios establecen que tanto las trayectorias como los posicionamientos de los agentes dentro de los campos no son iguales según se trate de mujeres u hombres. Entonces el investigador se propone construir un muestreo que responda a sus intereses teóricos. En este sentido, se plantea estudiar, por ejemplo, a ocho comunicadores, una mujer y un hombre, que se encuentren actualmente trabajando en el campo mediático; una mujer y un hombre que se desempeñen actualmente en el ámbito escolar; una mujer y un hombre que trabajen en el ámbito académico científico; una mujer y un hombre que trabajen en el ámbito estatal. Aquí lo que importa no es la cantidad de sujetos

ni generalizar los resultados de la muestra al total de la población. Lo importante es responder a preguntas teóricas. Se delimitan con antelación las variables de interés y los sujetos que se buscarán siguiendo esos criterios.

1.7.4.2.3 Muestreo bola de nieve

El muestreo *bola de nieve* (Taylor y Bogdan, 1987) es muy utilizado en estudios etnográficos pero, y al mismo tiempo, resulta una opción muy combinable con otras modalidades de investigación y con otros tipos de muestreo como el teórico, desarrollado fundamentalmente por Glaser y Strauss (1967), y el muestreo a criterio del investigador (Maxwell, 1996), visto en párrafos precedentes.

En los estudios etnográficos, resulta indispensable para los investigadores establecer buenas relaciones con una o dos personas dentro de la comunidad que se pretende estudiar. Estas personas se denominan *informantes* y generalmente facilitan el acceso al campo, y, al mismo tiempo, brindan información acerca de a quiénes más se puede entrevistar para ir ahondando en aquello que interesa conocer. Lo importante es, en palabras de Taylor y Bogdan (1987, 109) “conocer a algunos informantes y lograr que ellos nos presenten a otros”. Esta técnica utilizada tan frecuentemente en estudios etnográficos también puede traspolarse a estudios de tipo no etnográficos. Si pensamos en el ejemplo mencionado arriba, relacionado con el muestreo a criterio del investigador, podemos decir que, por ejemplo, una persona conocida y experta en la temática nos sugiera entrevistar en primer lugar a un hombre que ella conoce, que egresó en el año elegido para la investigación y que se encuentra hoy trabajando en un medio de comunicación. Tras entrevistar a este hombre, le preguntamos si no conoce a una mujer, de su misma cohorte, que esté también trabajando en el ámbito mediático y que pudiera estar dispuesta a ser entrevistada. Nos recomienda a una determinada persona. Entrevistamos a esta persona y ella nos sugiere a otra, y así sucesivamente.

1.7.4.2.4 Muestreo por cuotas

El muestreo por cuotas es el muestreo no probabilístico más utilizado en estudios de opinión o de mercado. Se trata de un caso especial del muestreo “a juicio del investigador” (Blanch y Joeques, 1997g).

Algunos autores, como Maintz, Holm y Hübner (1992), consideran a este muestreo como una técnica afín al muestreo aleatorio. Si el investigador conociera cómo se distribuyen todas las variables en una población, podría extraer muestras cuotificadas representativas del total. Sin embargo, esto resulta, en principio, imposible.

Para extraer un muestreo por cuotas es necesario, en primer término, conocer cómo se distribuye o caracteriza un reducido número de dimensiones (las de interés para el estudio) en el total de la población que se quiere estudiar. Esta información puede extraerse de censos, estadísticas oficiales o de listados completos (marcos de muestra) (Blanch y Joeques, 1997f; Maintz, Holm y Hübner, 1992).

Si se pretende realizar un estudio en el área de la Comunicación Social, con el objetivo de conocer las preferencias de consumo mediático televisivo de determinados subgrupos de la población, se puede diseñar un muestreo por cuotas, controlando determinadas características de los elementos del total de la población-objeto de estudio (como por ejemplo, tipo de empleo, sexo y edad). Lo primero será acceder a un censo de población e identificar la proporción de sujetos que se ubica en cada una de las intersecciones y cruces de las categorías de las variables de interés. Por ejemplo, las proporciones halladas (en la población total) pueden ser las siguientes (cuadro tomado de Maintz, Holm y Hübner, 1992, y adaptado):

	En relación de dependencia		Cuentapropista/ Pasante		Desempleado	
	Fem	Masc	Fem	Masc	Fem	Masc
Hasta 20 años	0%	0%	0%	0%	5%	5%
De 21 a 30 años	5%	5%	5%	5%	10%	10%
De 31 a 40	10%	10%	5%	5%	0%	0%
De 41 o más	10%	10%	0%	0%	0%	0%
Total	100%					

Si se quisiera extraer una muestra de 500 sujetos, su composición deberá corresponderse en números absolutos y relativos a los valores proporcionales indicados en la tabla de arriba (Maintz, Holm y Hübner, 1992).

Ahora bien, antes de hacer cualquier otro procedimiento habrá que determinar cuántos grupos de personas serán necesarios para realizar la encuesta. En el caso que analizamos, la variable “sexo” tiene 2 categorías; la variable “edad”, 4 y la variable “tipo de empleo”: 3. Multiplicamos $2 \times 4 \times 3 = 24$. El investigador necesitará 24 grupos de personas diferentes para poder controlar las tres características relevantes para su estudio. Los sujetos de cada muestra serán elegidos según los criterios de muestreo establecidos por el investigador, es decir, en función de las proporciones obtenidas a partir de la revisión de censos u otros documentos relevantes.

Si se quiere saber:

- cuántas mujeres de hasta 20 años desempleadas deberemos encuestar, habrá que multiplicar n (o tamaño de muestra, 500 sujetos) por la proporción (0,5 en este caso), es decir, $500 \times 0,5 = 25$
- cuántas mujeres de entre 31 y 40 años trabajan en relación de dependencia, tendremos que multiplicar $500 \times 0,10 = 50$, y así sucesivamente...

A continuación, un cuadro ilustrativo:

	En relación de dependencia		Cuentapropista/ Pasante		Desempleado	
	Fem	Masc	Fem	Masc	Fem	Masc
Hasta 20 años	0 pers	0 pers	0 pers	0 pers	25	25
De 21 a 30 años	25 pers	25 pers	25 pers	25 pers	50	50
De 31 a 40	50	50	25	25	0	0
De 41 o más	50	50	0	0	0	0
Total	500 personas					

Si el investigador, por ejemplo, tiene 5 encuestadores y decide que cada uno de ellos encueste a 100 sujetos, entonces, esos 100 sujetos elegidos por cada encuestador deberán respetar las proporciones poblacionales (expresadas en la tabla nº1).

En este muestreo el encuestador escoge a las personas a entrevistar en función de si éstas responden o no a la combinación de características dispuestas por el investigador (Maintz, Holm y Hübner, 1992). Por ello se sostiene que su dinámica es diferente a la del muestreo aleatorio. Mientras en el aleatorio, las muestras se extraen aleatoriamente de la población total, en el muestreo por cuotas, el investigador decide de manera arbitraria a quién o quiénes encuestar.

2. TÉCNICA DE ENCUESTA: EL CUESTIONARIO Y SU APLICACIÓN

2.1 Un poco de historia...

El sondeo o encuesta es un método científico de recolección de datos de carácter cuantitativo que permite recopilar información sobre opiniones, creencias y/o actitudes de los sujetos estudiados e indagar acerca de temas múltiples, tales como pautas de conducta o consumo, prejuicios sociales, trayectorias académicas, laborales, sociales, entre otros aspectos (Marradi, Archenti y Piovani, 2010). Brinda información acerca de cómo se manifiestan muchas propiedades (dimensiones, características o variables) en innumerables individuos y se aplica en ámbitos diversos tales como el comercial, el académico y el político (Marradi, Archenti y Piovani, 2010).

Basada “en la teoría estadística y las técnicas de muestreo” (Marradi, Archenti y Piovani, 2010, 179), los orígenes de esta técnica datan del siglo XIX. Sin embargo, sus primeras aplicaciones de envergadura (Marradi, Archenti y Piovani, 2010, 179), tuvieron lugar a partir de la década del 30 del siglo XX, en Estados Unidos, cuando George Gallup, psicólogo y periodista norteamericano, desarrolló uno de los más prestigiosos sondeos electorales que anticipó acertadamente el triunfo de Roosevelt en los comicios de 1936. Gallup, fundó a los 34 años, en 1935, el American Institute of Public Opinion; en 1936, el British Institute of Public Opinion y, en

1939, en Princeton, el Audience Research Institute. Los estudios de Gallup tuvieron como objeto a la opinión pública y el mercado, los mismos fueron requeridos por empresas cinematográficas tales como Walt Disney con el objetivo de conocer las características del público y adaptar sus ofertas a la demanda.

En el mundo académico de la Sociología y la Comunicación Social, el sondeo o encuesta, adquirió plena incorporación a partir de la década del 40. El austríaco Paul Lazarsfeld, desde el Departamento de Sociología de la Universidad de Columbia en Estados Unidos, enfocado en estudiar a los medios de comunicación y sus efectos en la conformación de la opinión pública, llevó a cabo “investigaciones sociales aplicadas basadas en encuestas” (Marradi, Archenti y Piovani, 2010, 179). Junto a Laswell, Lewin, Hovland y Merton (Mattelart, 1996; Wolf, 1996) dió inicio al desarrollo de la Sociología de la Comunicación de Masas. Luego, en la década del 50, hubo un incremento significativo de estudios académicos basados en sondeos.

La técnica de encuesta se convirtió prontamente en una de las más utilizadas en la investigación social empírica en los Estados Unidos, y se expandió y difundió hacia otros países desarrollados tales como Francia, Inglaterra, Alemania y Japón (Marradi, Archenti y Piovani, 2010). La encuesta resultó ser una de las herramientas “más potentes” (180) de los métodos cuantitativos, puesto que permitía obtener datos válidos en un plazo relativamente corto de tiempo y realizar inferencias y estimaciones (Marradi, Archenti y Piovani, 2010).

“La base estadística de la encuesta” (Marradi, Archenti y Piovani, 2010, 180) permite establecer predicciones representativas respecto de las características de grandes poblaciones a partir del análisis de los datos obtenidos en estudios a pequeña escala de “pequeñas muestras probabilísticas adecuadamente seleccionadas” (180). La encuesta permite describir cómo se manifiestan determinadas propiedades de una población, a partir del análisis

del comportamiento de las variables en una muestra, conociendo el margen de error. Siguiendo a Marradi, Archenti y Piovani (2010), como se trata de estudios probabilísticos, los resultados obtenidos son aproximaciones o estimaciones y no certezas. Ahora bien, no todos los sondeos responden a los muestreos probabilísticos, por lo tanto, no siempre se pueden estimar los parámetros de una población.

2.2 Algunos supuestos de la técnica de encuesta

Marradi, Archenti y Piovani (2010, 181) entienden, con Bourdieu (1973), que algunos supuestos *inverificables* de las encuestas de opinión son los siguientes:

Por una parte, todos los sujetos tienen opinión propia; por otra, pueden cuestionarse internamente acerca de los temas sobre los que se les pregunta; finalmente, todas las opiniones tienen una misma incidencia social

Los dos primeros aspectos mencionados arriba tienen relación con las implicancias de una situación de encuesta (Marradi, Archenti y Piovani, 2010). La persona encuestada puede sentirse incómoda y obligada a responder, con temor a que su respuesta sea evaluada negativamente por el encuestador. Esto deriva directamente en el problema de la veracidad de las respuestas. Como se trata de un “problema irresoluble” (Marradi, Archenti y Piovani, 2010, 181), entonces se tiende a tomar a la opinión del sujeto como si fuera siempre verdadera.

El último aspecto mencionado, tiene relación con el argumento acerca de que en investigaciones centradas en grandes poblaciones, cada individuo es tomado como si fuera equivalente a los demás, resultando cada uno de ellos sustituible por otro (Marradi, Archenti y Piovani, 2010, 181). Este principio permite trabajar las opiniones de los individuos “matemáticamente, bajo

el supuesto de la acumulatividad de las respuestas” (182) aunque perdiendo, simultáneamente, información específica relativa a cada caso particular.

2.3 Tipos de estudios

2.3.1 Estudios comparados

Se centran en aplicar cuestionarios en diferentes culturas o países, replicando la encuesta de manera sistematizada a intervalos regulares en el tiempo (Marradi, Archenti y Piovani, 2010). Son estudios de gran alcance que permiten construir importantes bases y bancos de datos. Un ejemplo es la International Social Survey Program (ISSP), el Eurobarómetro y el Latinobarómetro (Marradi, Archenti y Piovani, 2010). El Latinobarómetro, es una ONG con sede en Santiago de Chile que se dedica a estudiar la opinión pública en dieciocho países latinoamericanos.

El problema principal de estos estudios es la comparabilidad de los datos, puesto que recolectan información referida a países y realidades sociales, económicas y culturales muy diversos/as (Marradi, Archenti y Piovani, 2010).

2.3.2 Estudios longitudinales

Los *estudios longitudinales o de panel* requieren de la aplicación de un cuestionario “a una misma muestra en diferentes momentos a lo largo del tiempo” (Marradi, Archenti y Piovani, 2010, 186). Son estudios que permiten conocer las variaciones de las posiciones sociales de los sujetos en el tiempo, los cambios de actitudes, opiniones y conductas a partir de la acción de determinados factores sean éstos macro o microestructurales (o sociales), endógenos o exógenos; permiten, además, realizar comparaciones entre generaciones, cohortes, entre otros.

2.4 ¿Qué es y para qué sirve la encuesta?

Una vez que se decide el tipo de muestreo a utilizar en la investigación, resulta necesario armar el instrumento de recolección de información. El instrumento de medición es el cuestionario.

Una definición técnica de la encuesta, la define como: una herramienta o instrumento estandarizada/o que permite obtener información acerca de una muestra de la población total (Dorado, 2002). Por su parte, Campbell y Katona, (1979, 31), señalan: “Es una técnica que depende del contacto directo o indirecto con todas las personas cuyas actitudes, conductas o características son significativas para una determinada investigación”.

Particularmente, el cuestionario permite obtener información sobre las personas en estudio de manera sistemática y ordenada. Esa información puede ser sobre sus trayectorias académicas, laborales, sociales, familiares, vitales; sobre lo que ellos hacen, opinan, esperan, quieren o desprecian, aprueban o desaprueban; sobre los motivos de sus actos, opiniones y actitudes; sobre las características personales de los individuos (como su edad, sexo, estado civil, grado de instrucción); las características estructurales o medio ambiente dentro del cual aquéllos están insertos, entre otros (Sierra Bravo, 1988).

2.5 Cuestiones éticas

Para realizar una encuesta son necesarios los preceptos éticos. Según Sierra Bravo (1988) el sujeto que se estudia debe conocer de antemano los objetivos del estudio y debe dar un consentimiento explícito a la situación de encuesta. Si esto no ocurre, el investigador no tiene derecho a inmiscuirse en la privacidad de las personas. Por otra parte, el investigador deberá: a) asegurar al *respondente* la confidencialidad de la información por él brindada, b) realizar un tratamiento confidencial de los datos una vez concluida la instancia de encuesta. El investigador no debe dar a conocer los nombres de

las personas encuestadas bajo ninguna circunstancia. El uso que haga de los datos recabados debe responder a los fines conocidos y autorizados por los respondientes.

2.6 Cuestiones generales de procedimiento

Sierra Bravo (1988, 323) propone llevar a cabo las siguientes acciones:

- 1) Transmitir a los encuestados con claridad los objetivos de la encuesta, el nombre de la entidad que la propicia, el modo de uso de los datos (asegurando la confidencialidad de los mismos).
- 2) Dar respuestas precisas y dedicadas a las preguntas y dudas de los sujetos en relación con el contenido de las mismas y la manera de responderlas.
- 3) Asegurar a los sujetos la posibilidad de participar o no en la encuesta y de abandonarla en caso de que así lo deseen.
- 4) Hacer de los datos obtenidos sólo el uso para el que estaban previstos, manteniendo siempre el anonimato de los informantes.

2.7 Construcción del cuestionario

La construcción del instrumento de medición se lleva a cabo en varias etapas.

2.7.1 Determinación de variables, nivel de medición y operacionalización

En primer lugar, debemos identificar las variables que queremos medir. Para ello debemos primero, definir conceptualmente cada variable, reconocer su nivel de medición, identificar las dimensiones de cada una y especificar los indicadores de cada dimensión.

Este paso es fundamental porque contribuye a determinar los datos o puntos de información que vamos a obtener mediante la encuesta, los cuales quedarán manifiestos a partir de las preguntas presentes en el cuestionario. En palabras de Bravo (1988, 316), estos puntos de información deben ser sometidos a “juicio crítico” para determinar si se incluyen o no en el cuestionario, y completados “con aquellos aspectos que si bien no derivaron de la operacionalización de variables fueron considerados convenientes por el interés que revisten. Además, cuando existe la posibilidad, resulta conveniente revisar la manera en que otros investigadores abocados a la temática de interés construyeron sus propios cuestionarios. Esto permite comparar el cuestionario propio con los realizados por terceros, suplir falencias propias a partir de la adopción y/o adaptación de aquello realizado por otros; pero, y al mismo tiempo, esto también permite aprender a no repetir los errores cometidos por otros (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000).

2.7.2 Tipo, cantidad, disposición de las preguntas y su codificación

La segunda etapa en la ideación de cuestionarios consiste en (Sierra Bravo, 1988:, 316):

- a) Especificar las preguntas, según sean las variables a medir y los tipos de preguntas que permitan obtener los “datos pertinentes” (Quivy y Compenhoudt, 1998) establecidos por los indicadores de las dimensiones de las variables (es decir, hay que determinar los tipos de preguntas que se van a plantear, si preguntas abiertas o cerradas y las categorías u opciones de respuesta más convenientes).

- b) Especificar el número, orden y disposición de las preguntas en el cuestionario (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000, 278);

- c) Determinar la forma de aplicación del cuestionario (si por escrito, mediante entrevista, a través de Internet);
- d) Codificar cada una de las preguntas y sus respectivas categorías de respuesta (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000). Por ejemplo, si la primer pregunta de un cuestionario es: “¿Lee ud. diarios?”, y las respuestas posibles: “sí / no”, entonces, delante de la pregunta colocamos el número 1; delante de la respuesta “sí” colocamos el número 1 y delante del “no”, el número 2. De este modo, queda codificada la pregunta con sus respectivos indicadores.

2.8 Prueba piloto del instrumento

La prueba piloto del instrumento de medición (o cuestionario) consiste en aplicarlo a una pequeña muestra del total de la población (Scheaffer, Mendenhall y Ott, 1987; Sampieri, Collado y Lucio, 2000). El objetivo es determinar si las instrucciones de llenado de la planilla son claras, si las preguntas formuladas son comprendidas por los encuestados, si éstas responden a lo que queremos estudiar, entre otros aspectos. Encuestar a un grupo reducido de personas permite según Scheaffer, Mendenhall y Ott (1987, 35-36) “probar el cuestionario y verificar el manejo de las acciones que debe realizar el encuestador en el campo”. Los resultados obtenidos permiten modificar el cuestionario antes de realizar el muestreo a gran escala.

La aplicación temprana del cuestionario permite su modificación, a partir de la eliminación de preguntas innecesarias o confusas, o de aquellas que brindan respuestas repetitivas; de la remodelación de algunas, y la inclusión de otras nuevas (tomando en consideración lo manifestado por los encuestados). Las respuestas de los sujetos estudiados permiten no solo eliminar opciones de respuesta innecesarias sino, además, incorporar ciertas otras categorías

o indicadores impensados durante el proceso de construcción del instrumento. Así es como se incorporan a la cédula definitiva aspectos no previstos inicialmente pero que son relevantes para el estudio. Por otra parte, si, por caso, un mismo cuestionario debiera ser aplicado a grupos de sujetos distantes entre sí, cada uno de ellos situados en contextos sociales, culturales, económicos y políticos diferenciados, será necesario probar el instrumento en cada una de las comunidades por separado antes de realizar el muestreo a gran escala.

La prueba piloto es fundamental para evaluar si el instrumento recaba la información necesaria para la investigación. Es decir, a partir de este ensayo, y en términos cuantitativos, el investigador evalúa si el instrumento es válido, esto es, si mide lo que inicialmente pretendía medir (Maintz, Hölm y Hübner, 1992).

2.9 Recomendaciones para armar un cuestionario

2.9.1 Presentación del cuestionario

Todo cuestionario debe ser presentado ante el sujeto encuestado. Algunos aspectos importantes a tener en cuenta en este sentido son los siguientes:

- a) Presentación y propósitos del cuestionario al mismo tiempo que la aclaración respecto del método de selección de personas y del uso que se hará de la información recabada (confidencialidad de la información).
- b) Instrucciones claras y concisas acerca de cómo responder el cuestionario.
- c) Agradecimiento al encuestado por haber colaborado con la encuesta.

A continuación se presentan a modo de ilustración tres ejemplos de presentación de cuestionarios. Dos han sido adaptados de Hernández Sampieri, Fernández Collado y Baptista Lucio (2000).

El tercero, se corresponde con una presentación realizada por una organización no gubernamental.

Un ejemplo tomado y adaptado de Hernández Sampieri, Fernández Collado y Baptista Lucio (2000, 287) es el siguiente:

Buen día:

Estamos realizando un estudio con el objetivo de.....

Tenemos interés en que contestes una serie de preguntas que no te tomarán mucho tiempo. Tus respuestas serán confidenciales y anónimas.

Las personas seleccionadas para el estudio se eligieron al azar. Las opiniones recabadas serán utilizadas en....., y nunca se comunicarán datos individuales.

Te agradecemos responder con la mayor sinceridad posible. No hay respuestas correctas ni incorrectas.

Sugerimos leer las preguntas con detenimiento. En algunas, sólo puedes seleccionar una opción; en otras, puedes elegir varias opciones y en unas pocas, puedes responder libremente.

Muchas gracias por tu colaboración.

Otro ejemplo, también tomado de Hernández Sampieri, Fernández Collado y Baptista Lucio (2000, 287) es el que se presenta a continuación:

Buenos días (tardes).

El área de comunicación de Spa Sociedad Anónima está haciendo una encuesta para conocer la opinión de los clientes respecto de nuestros servicios. Es por ello que lo invitamos a contestar unas preguntas. No le tomará más de 20 minutos. La información que nos proporcione será manejada con la más estricta confidencialidad. No hay preguntas comprometedoras.

Un tercer ejemplo podría ser:

Tenemos interés en conocer su opinión respecto de nuestras actividades. Para ello lo invitamos a responder el siguiente cuestionario marcando con una cruz la opción que mejor se adecue a su percepción y expresando sus comentarios si lo desea. La información será tratada con confidencialidad. Le agradecemos su colaboración.

2.9.2 Recomendaciones para redactar buenas preguntas

Veamos algunas recomendaciones para formular las preguntas.

- Para Cannell y Kahn (1979), las preguntas del cuestionario deben tener una progresión lógica (empezando por preguntas simples hasta llegar a las más complejas) e ir introduciendo al informante de manera paulatina en la situación de encuesta. Por su parte, Padua y Ahman (1979) recomienda formular, en primer término, las preguntas más generales para plantear luego las más restringidas o particulares. Los autores plantean, además, evitar requerirle a los sujetos datos personales de manera prematura.

- Sierra Bravo (1988, 316) agrega a lo anterior que todas las preguntas referidas a un mismo tema o aspecto deberían ir juntas en el cuestionario, guardando un orden temporal, lógico y psicológico. En el mismo sentido, Cannell y Kahn (1979) sugieren además, conducir el interrogatorio de manera tal que no se realicen saltos bruscos entre un marco de referencia temático y otro. Bravo (1988) agrega, además que, para evitar el cansancio por parte del encuestado, es recomendable intercalar en sitios estratégicos preguntas más o menos intrascendentes, que le permitan al encuestado tomarse un respiro.
- En palabras de Cannel y Kahn (1979, 326) las preguntas del cuestionario no deben constituir nunca una amenaza para el sujeto. Si, por ejemplo, en una institución queremos conocer el grado de aceptación que los empleados tienen respecto de sus superiores, y preguntamos: “¿qué opinás de tu superior?”, el sujeto, muy probablemente responda según lo que considere sea una contestación socialmente aceptable en el medio en el que trabaja. El individuo difícilmente reconozca ante el encuestador que, por ejemplo, está disconforme con su superior.
- Un error frecuente en los cuestionarios autoadministrados es el denominado “error de pericia” (Cannel y Kahn, 1979, 325). Consiste en el error de adscribir al sujeto un alto grado de conocimiento sobre determinada temática cuando éste no necesariamente lo posee. Las preguntas perito o especializadas pueden exigir al sujeto que se manifieste en relación con algún aspecto de la realidad sobre el cual nunca antes fue consciente. Esto puede incomodar y producir en el encuestado una desmotivación para comunicarse con el encuestador. Por ejemplo, supongamos que en una institución el equipo de comunicación institucional plantea, en un cuestionario, una pregunta como la siguiente: “¿Cuáles son los beneficios para la cooperativa de mantener una

misión, una visión y objetivos particulares?” En este caso, el encuestado no experto probablemente se sienta descalificado e incapaz para responder a esa pregunta.

- Las preguntas deben ser formuladas con cautela (Cannel y Kahn, 1979, 325). Lo anterior debido a que puede haber un amplio espectro de respuestas y porque no es aconsejable incomodar a nadie. En el ejemplo antedicho, si los miembros del equipo de comunicación tuvieran en cuenta que no todo el personal sabe lo que significa o implica la misión, visión y objetivos de una empresa, y siguiendo los lineamientos de Cannel y Kahn (1979), quizás podrían reformular la pregunta del siguiente modo: “No todas las personas han tenido la oportunidad de conocer la utilidad de la misión, la visión y objetivos particulares de una institución pero algunos reunieron información sobre ese tema. Acaso ¿sabe usted cuáles son los beneficios de que la institución mantenga una misión, una visión y objetivos?”.
- En la misma línea de razonamiento, podemos también citar a Hernández Sampieri, Fernández Collado y Baptista Lucio (2000, 284), quienes plantean la imperiosa necesidad de no incomodar a los sujetos estudiados. Si, por ejemplo, un investigador está estudiando el consumo de alcohol en un subgrupo de la población, en vez de preguntar de manera directa: “¿acostumbra consumir de bebida alcohólica?”, quizás resulte más conveniente empezar con preguntas generales como: “¿algunos de sus amigos acostumbra consumir bebidas alcohólicas?”, para ir luego afinando las preguntas e inquirir de manera sutil, por ejemplo, “¿cuál es su tipo de bebida favorita?”.
- Las preguntas deben ser claras y comprensibles (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000). Si quisiéramos estudiar el nivel de consumo diario de televisión

de un poblado x, en vez de preguntar “¿ve usted televisión?”, pregunta vaga e imprecisa, podríamos formular una serie de preguntas más específicas y claras. Por ejemplo: “¿Acostumbra usted ver televisión diariamente?; ¿cuántos días durante la última semana vio televisión?; ¿en qué horarios del día usted ve televisión?; ¿qué canales ve más en la semana?; ¿qué programas ve con mayor frecuencia?” (284), etc.

- El lenguaje a utilizar en los cuestionarios debe ser acorde con las características socioeconómicas y demográficas de las personas que se van a encuestar (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000). Esto es, antes de redactar las preguntas resulta imperioso tomar en consideración la edad, la condición social, el nivel social y educativo y la capacidad de rememoración de hechos pasados de los sujetos.
- Según Cannel y Kahn (1979, 327) y Hernández Sampieri, Fernández Collado y Baptista Lucio (2000), las preguntas deben limitarse a una única referencia o a un solo aspecto o relación lógica. Por ejemplo, si un investigador quiere determinar cuál es la actitud de determinados sujetos hacia los planes *jefas y jefes de hogar* y hacia el seguro de desempleo, no resultará productivo preguntar: “¿Cuál es su opinión (o si está a favor o en contra, si se trata de pregunta con opciones cerradas de respuestas) respecto del seguro de desempleo y los planes jefes y jefas de hogar?”. Lo ideal sería formular dos preguntas separadas, cada una de ellas referida a uno de los dos aspectos. Del mismo modo, para un investigador que está estudiando el consumo de medios masivos de comunicación, no sería conveniente formular una pregunta como la siguiente: “¿Acostumbra usted ver televisión y escuchar radio diariamente?”. En todo caso podría plantear, al igual que en el ejemplo anterior, dos preguntas separadas.

- Un aspecto fundamental en la formulación de preguntas es que las mismas no induzcan las respuestas (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000). Es común encontrar en cuestionarios preguntas tendenciosas que orientan al sujeto a elegir un tipo de respuesta en detrimento de otras. Es el caso de preguntas como: “¿Considera a esta empresa *mejor* posicionada que ...?. ¿Los planes de capacitación en las empresas en XY ciudad son *muy eficaces*? ¿La manera en que el diario *La Nación* presenta el tema xx es parcial?”. En los tres casos, además de tratarse de juicios de valor, se induce al sujeto a responder afirmativamente a la pregunta. En contraste, estas preguntas bien formuladas podrían expresarse como sigue: “¿Cómo posicionaría usted a esta empresa en el mercado?” o “¿Cómo considera usted que se posiciona esta empresa en el mercado?”; “¿Cómo definiría/calificaría usted a los planes de capacitación de las empresas de XY ciudad” o “¿Cómo considera usted que son los planes de capacitación de las empresas de XY ciudad?”; “¿Cómo considera usted que es tratado el tema xx por el diario *La Nación*?”.
- Es recomendable que las preguntas no se respalden en instituciones, ideas, personalidades reconocidas ni en evidencia comprobada (Sampieri, Collado y Lucio, 2000). Por ejemplo, un investigador médico está haciendo un estudio sobre tabaquismo y formula una pregunta como la que sigue: “La Organización Mundial de la Salud ha realizado diversos estudios y concluyó que el tabaquismo produce daños al organismo, ¿usted considera que fumar es nocivo para la salud” (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000, 285). Se trata de una pregunta tendenciosa, parcial. Lo mismo cabe para preguntas como: “La mayoría de las personas piensa que...” o “La Iglesia considera que...”; “La Facultad de Comunicación de la X Universidad ha adoptado xx postura ante la Ley de Radiodifusión. ¿Qué opinión le merece a usted la Ley de Radiodifusión?” Un comienzo de este

tipo puede condicionar la respuesta del encuestado. Este tipo de preguntas es acorde sólo estudios de actitud. En estudios de actitud se escriben afirmaciones ante las cuáles el sujeto debe establecer su grado de acuerdo o desacuerdo.

- Cuando hay varias alternativas de respuesta y el encuestado debe elegir una, existe la posibilidad de que el orden de las categorías afecte sus elecciones (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000). Resulta conveniente, entonces, “rotar el orden de lectura de las alternativas” (285). Por ejemplo, si un investigador de mercado quiere conocer las preferencias de consumo de un amplio grupo de sujetos respecto de jabones en polvo para lavarropas, probablemente tendrá que intercalar las marcas de los mismos (mezclando primeras con segundas marcas e incluso con marcas apenas conocidas en el mercado). El investigador deberá presentar las opciones de respuesta de manera intercalada para evitar que las personas marquen con una cruz sólo en los casilleros correspondientes, por caso, con las primeras marcas. Por ejemplo: “¿Qué jabón en polvo prefiere usted...?” Opciones de respuesta: “Zorro, Gigante, Skip, Ala, Drive...”, etc... Importante es aclarar aquí que para las escalas de Lickert esta regla no se aplica.

2.9.3 Tipo de preguntas

2.9.3.1 La pregunta cerrada

La pregunta cerrada implica, para el respondente, tener que elegir entre una u otra categoría de respuesta, según sea su posición. Este tipo de pregunta resulta útil en situaciones en las cuales:

- a) existe un solo marco de referencia a partir del cual la pregunta puede ser contestada por el sujeto;
- b) dentro de este marco (de referencia) existe una gama conocida de respuestas posibles;
- c) dentro de esta gama

existen puntos de elección claramente definidos, los cuales representan con precisión la posición de cada sujeto (Cannell y Kahn, 1979, 330).

Por ejemplo, la pregunta “¿Es usted?” permite clasificar a los individuos según su *status* conyugal actual a partir de una gama de posibles respuestas: soltero, casado, divorciado, separado, viudo, juntado. Las mismas son claras y hay un solo marco de referencia para todos los sujetos.

2.9.3.2 La pregunta cerrada simple

Es cualquier pregunta para la cual existen, mayoritariamente, dos o tres alternativas de respuestas posibles (Padua y Ahman, 1979).

Un ejemplo puede ser (99):

¿Trabaja ud. actualmente?	Sí <input type="checkbox"/> No <input type="checkbox"/>
---------------------------	--

Son preguntas fáciles de registrar, interpretar, codificar y analizar, y no requieren de encuestadores altamente calificados. Una desventaja es la imposibilidad de realizar clasificaciones detalladas. Según Padua y Ahman (1979) aquellos sujetos cuya situación no corresponda ni con la opción de respuesta 1 ni con la 2, deberán responder forzosamente en uno u otro sentido. Para resolver esta dificultad el autor sugiere formular estas preguntas haciendo explícitas las instrucciones de llenado del formulario. Por ejemplo, en el caso de la pregunta planteada arriba, el encuestador podría aclarar qué se entiende por trabajo, para, en caso de encontrarse con un jubilado que como *hobby* trabaje 10 horas semanales, sepa en qué categoría debe ser ubicado. Lo mismo si se encontrara con alguien que esté trabajando temporariamente al momento de la entrevista.

2.9.3.3 La pregunta cerrada con múltiples respuestas

La pregunta cerrada con múltiples respuestas suele denominarse “pregunta de cafetería” puesto que la misma contiene

varias o múltiples alternativas de respuesta, y, en consecuencia, código múltiple. Se trata de un tipo de preguntas que requiere de encuestadores preferentemente entrenados. La presencia de varias alternativas de respuesta, permite recoger mayor cantidad de información. Ahora bien, para obtener mayor información, resulta necesario obtener, previamente, las categorías de respuesta adecuadas. Para saber si las posibles alternativas de respuesta son correctas, Padua y Ahman (1979: 99-100) sugiere realizar un pre-test o prueba piloto del cuestionario.

Veamos el ejemplo (tomado de Padua y Ahman, 1979, 100, y adaptado):

<p>1. ¿A qué hora del día parte ud. a su trabajo? Indique la hora con precisión horaria.</p>	<p><input type="checkbox"/> 1. Los lunes parto de mi casa a las.... Los martes a las ... Los miércoles a las... Los jueves a las ... Los viernes, a las..... Los sábados, a las..... Los domingos, a las....</p> <p><input type="checkbox"/> 2. Mi horario de salida no sigue los días de la semana (pase a la pregunta número X)</p> <p><input type="checkbox"/> 3. No parto para mi trabajo a una hora especial El entrevistado es..... </p> <p><input type="checkbox"/> 4. Realizo mi trabajo en el hogar El entrevistado es..... </p> <p><input type="checkbox"/> 5. No tengo ninguna profesión El entrevistado es..... </p> <p><input type="checkbox"/> 6. Otro:.....</p>
--	---

2.9.3.4 Preguntas cerradas con más de una opción para elegir (con categorías excluyentes y no mutuamente excluyentes)

Las preguntas cerradas con más de una opción para elegir, mantienen opciones de respuestas no excluyentes (o, lo que es lo mismo, no mutuamente excluyentes). Contrariamente, las preguntas cerradas con una sola opción para elegir, mantienen opciones de respuesta mutuamente excluyentes. A continuación dos ejemplos tomados de Hernández Sampieri, Fernández Collado y Baptista Lucio (1998, 278) y Hayes (1998, 160), respectivamente.

Los medios utilizados para la comunicación dentro de la escuela son:
(Marque con una cruz las opciones que correspondan)

- | | |
|---------------------------|--------------------------|
| teléfono | <input type="checkbox"/> |
| mail | <input type="checkbox"/> |
| paneles | <input type="checkbox"/> |
| memos | <input type="checkbox"/> |
| Comunicación cara a cara | <input type="checkbox"/> |
| Ninguno de los anteriores | <input type="checkbox"/> |

Por favor, seleccione la categoría que mejor describa su tipo de empleo/ ocupación (seleccione sólo una opción)

- | | |
|--|--|
| <input type="checkbox"/> educador | <input type="checkbox"/> trabajo en construcción
(como mano de obra calificada:
ingenieros, arquitectos) |
| <input type="checkbox"/> trabajo en salud | <input type="checkbox"/> estudiante |
| <input type="checkbox"/> ama de casa | <input type="checkbox"/> ejecutivo / management |
| <input type="checkbox"/> desocupado | <input type="checkbox"/> empleado público/
Administrativo |
| <input type="checkbox"/> empresario | <input type="checkbox"/> funcionario |
| <input type="checkbox"/> comerciante | <input type="checkbox"/> profesión liberal |
| <input type="checkbox"/> jubilado | <input type="checkbox"/> no contesta |
| <input type="checkbox"/> otros. Especifique: _____ | |

2.9.3.5 Preguntas cerradas de jerarquización de opciones

Veamos un ejemplo tomado y reformulado, en parte, de Berenson y Levine (1996, 28):

¿Qué característica de su trabajo es la más importante para usted? ¿Cuál ubicaría en un segundo lugar? ¿Cuál en tercer lugar? Coloque el número 1 al lado de la característica que considere más importante y así sucesivamente hasta llegar a la tercera.

- | | |
|--|---|
| <input type="checkbox"/> Altos Ingresos | <input type="checkbox"/> Disfrute del trabajo |
| <input type="checkbox"/> Oportunidades de progreso | |
| <input type="checkbox"/> Horario flexible | <input type="checkbox"/> Ningún riesgo de ser despedido |

2.9.3.6 Preguntas abiertas

La pregunta abierta estructura un tema para el sujeto, aunque le deja a su libre arbitrio la tarea de responder, haciendo uso de sus propias palabras, según lo que le parezca conveniente (Cannell y Kahn, 1979, 330).

2.9.3.6.1 Tratamiento de las respuestas a preguntas abiertas

En este tipo de preguntas las alternativas de respuestas son infinitas. Una ventaja en relación con estas preguntas es la posibilidad de comprender cuáles son los por qué de las conductas, decisiones y actitudes de los sujetos. Los agentes encuestados pueden expresar y expresar su propia subjetividad, la manera en que ellos perciben las situaciones en las que están involucrados. Aquí ellos pueden expresarse con sus propias palabras, sus motivos, sus razones, deseos, modos de pensar, entre otros aspectos. Padua y Ahman (1979, 103) sostienen que se trata de preguntas que incrementan la motivación del encuestado durante el tiempo que dura la encuesta.

Una de las desventajas es la dificultad para responder por parte de algunos *respondentes* puesto que sienten que deben improvisar una respuesta rápida sin preparación previa.

Por otra parte, otra desventaja es que las respuestas abiertas brindadas por los sujetos mantienen cierta dificultad de interpretación. El investigador debe realizar un esfuerzo extra para comprender lo que los encuestados han querido expresar.

De una buena interpretación de las respuestas depende la posterior reducción de los datos a categorías y códigos para su posterior procesamiento cuantitativo/estadístico. Un método de trabajo con respuestas a preguntas abiertas es el método de *codificación en el campo* (Padua y Ahman, 1979). El entrevistador, a partir de la respuesta espontánea del sujeto encuestado, debe estimar sus sentires y pensamientos, y codificarlos, cuasi-automáticamente en categorías pre-establecidas. Padua y Ahman (1979) advierten que, cuando se aplica este método, existe siempre la posibilidad de que el encuestador introduzca su propio sesgo subjetivo al momento de estimar lo que el sujeto encuestado quiso expresar. Sus propios juicios de valor pueden producir distorsiones en la interpretación que hace de lo que el otro piensa, cree, siente, expresa. Para este tipo de preguntas generalmente se requiere de encuestadores bien preparados.

Una forma más interesante de trabajar las respuestas a preguntas abiertas es a partir de la utilización del método de *categorización temática* y/o *categorización in vivo*. El primero, responde a una lógica deductiva de análisis. El segundo a una lógica inductiva. Ambas pueden estar presentes en el análisis de datos cualitativos. El resultado de ese análisis (obtener ciertas categorías) puede ser utilizado para realizar luego un tratamiento cuantitativo de datos.

La categorización temática, utilizada mayormente en estudios cualitativos, puede servir para trabajar las respuestas a preguntas abiertas de un cuestionario. Como primer paso, deberían codificarse los párrafos de la encuesta, asignándole a cada uno número. Cada párrafo a su vez, debe segmentarse para encontrar en él los elementos singulares. Siguiendo a Rodríguez Gómez, Gil Flores y García Jiménez (1996, 206), “del conjunto global de datos pueden diferenciarse segmentos o unidades de observación que pueden resultar relevantes y significativas para la investigación”.

Para identificar y clasificar las unidades de observación se procede según el criterio de categorización temática. La categorización, según Rodríguez Gómez, Gil Flores y García Jiménez (1996) permite clasificar conceptualmente a las unidades de observación que se engloban dentro de un mismo tópico. Una categoría, cualquiera sea, mantiene un significado o tipo de significados y “puede referirse a situaciones y contextos, actividades y acontecimientos, relaciones entre personas, comportamientos, opiniones, sentimientos, perspectivas sobre un problema, métodos y estrategias, procesos” (p.208). Este método permite entonces identificar unidades de observación a partir de categorías predeterminadas y establecidas por el marco teórico conceptual elegido en la investigación o por los conceptos o variables de las hipótesis que han sido definidos y operacionalizados al comienzo del trabajo de campo.

El segundo modo de proceder para el análisis de los datos, la codificación abierta, responde al modo de operar de la Grounded Theory. Strauss y Corbin (1990; 1994); Glasser y Strauss (1967); Strauss (1989) entienden que la categorización puede realizarse durante el mismo proceso de recolección de datos, lo cual permite construir categorías provisionales o *in vivo* que servirán luego para generar otras más abarcativas (categorías englobantes) que incluyan a varias de las denominadas *in vivo*. Las categorías *in vivo* son aquellas que surgen de boca de los entrevistados. En la medida en que se avanza en la codificación éstas se pueden ir modificando, consolidando o suprimiendo a partir del método de comparación constante aplicado a los datos reunidos bajo una misma categoría englobante o a partir de la comparación de datos de una categoría con datos de otra. Este procedimiento responde a una lógica inductiva.

Además de todo lo anterior, también puede optarse por transcribir textualmente al final del informe estadístico las respuestas de los encuestados tal cual ellos las manifestaron. Todo dependerá de las necesidades y requerimientos de la investigación que se esté llevando a cabo.

2.9.3.7 Utilización de preguntas cerradas y abiertas, en conjunto

Un ejemplo podría ser el que Padua y Ahman (1979, 106-107) proponen como modelo para estudiar la exposición y percepción de medios de comunicación de masas, en particular de los diarios y revistas de un país determinado. Para ello, los autores recomiendan empezar preguntando primero si la persona lee o no el diario, luego, en caso de que lea el diario, qué diario lee; a continuación, qué lee el entrevistado dentro del diario, y finalmente con qué frecuencia lee ese diario. Padua y Ahman (1979) sugieren seguir un orden lógico en la formulación de las preguntas que irán en el cuestionario, ir de lo más general a lo más particular, de lo más simple a lo más complejo, esto es, ordenar los diferentes campos de preguntas de acuerdo a la manera lógica y natural de pensar, primero en el área más grande (si lee o no los diarios), preguntando luego por campos más restringidos (qué partes del diario lee), y así sucesivamente. Para ello, se combinan preguntas cerradas simples, cerradas con respuestas múltiples y abiertas. A continuación un ejemplo tomado y adaptado de Padua y Ahman (1979, 106-107):

10. ¿Lee ud. diarios?	1. <input type="checkbox"/> Sí 2. <input type="checkbox"/> No (Pasar a la pregunta núm. 8)
11. ¿Cuál es el diario de tirada nacional que ud. lee con mayor frecuencia?	1. <input type="checkbox"/> Clarín 2. <input type="checkbox"/> La Nación 3. <input type="checkbox"/> Página 12 4. <input type="checkbox"/> La Prensa 5. <input type="checkbox"/> Ámbito Financiero 6. <input type="checkbox"/> Otro ¿Cuál?.....
12. ¿Usted lee también otro diario? ¿Cuál? Señalar el que lee en 2do. lugar.	0. <input type="checkbox"/> No lee otro diario 1. <input type="checkbox"/> Clarín 2. <input type="checkbox"/> La Nación 3. <input type="checkbox"/> Página 12 4. <input type="checkbox"/> La Prensa 5. <input type="checkbox"/> Ámbito Financiero 6. <input type="checkbox"/> Otro ¿Cuál?.....

<p>13. Ud. ha dicho que lee el..... (Mencionar aquí el diario que lee en primer lugar).</p> <p>Lo hace (marcar alguna de las opciones de respuesta en el cuadro a la derecha):</p>	<p>1. <input type="checkbox"/> Todos los días (5-7 días de la semana)</p> <p>2. <input type="checkbox"/> Algunas veces por semana (2-4)</p> <p>3. <input type="checkbox"/> Una vez por semana (1)</p> <p>4. <input type="checkbox"/> Algunas veces al mes (1-3)</p> <p>5. <input type="checkbox"/> Casi nunca</p> <p>6. <input type="checkbox"/> Otro</p> <p>¿Cuál?.....</p>
<p>14. ¿Qué partes del diario que lee con mayor frecuencia le gusta leer más? Se hace referencia aquí al diario citado en las preguntas 11 y 13.</p>	
<p>15. Ud., ha dicho que también lee el..... (mencionar aquí el diario que lee en segundo lugar).</p> <p>Lo hace (marcar alguna de las opciones de respuesta en el cuadro a la derecha):</p>	<p>1. <input type="checkbox"/> Todos los días (5-7 días de la semana)</p> <p>2. <input type="checkbox"/> Algunas veces por semana (2-4)</p> <p>3. <input type="checkbox"/> Una vez por semana (1)</p> <p>4. <input type="checkbox"/> Algunas veces al mes (1-3)</p> <p>5. <input type="checkbox"/> Casi nunca</p> <p>6. <input type="checkbox"/> Otro</p> <p>¿Cuál?.....</p>
<p>16. ¿Qué partes en particular le gusta leer en el.....? Mencionar el segundo diario.</p>	
<p>17. ¿Ud. lee revistas?</p>	<p>1. <input type="checkbox"/> Sí</p> <p>2. <input type="checkbox"/> No</p>

2.9.3.8 Embudo de preguntas

Cannell y Kahn (1979) sostienen que el enfoque *embudo* (visto ya en el ejemplo anterior) se refiere al procedimiento de plantear primero las preguntas más generales o menos restringidas para seguir luego con los aspectos más limitados. Según estos autores, esta secuencia de preguntas contribuye a determinar el marco de referencia de los sujetos investigados. Veamos el siguiente ejemplo (Cannell y Kahn, 1979, 328):

1) En general, ¿cómo describiría las relaciones en su lugar de trabajo?

Muy buenas Buenas Regulares Malas Muy malas

2) ¿Cómo considera usted que son las relaciones laborales entre la gerencia y los empleados?

Muy buenas Buenas Regulares Malas Muy malas

3) ¿Piensa usted que la gerencia debería establecer una forma diferente de relación con los empleados a la actual?

sí no

4) (Si contestó afirmativamente) ¿Qué considera usted que habría que modificar?

5) Algunas personas dicen que las relaciones entre gerentes y empleados son muy distantes y otros que son muy cordiales ¿cuál es su opinión al respecto?

Como puede observarse, la primera pregunta tiene un enfoque muy general. No establece ninguna tendencia de pensamiento respecto de un sector específico de la empresa. En la segunda pregunta se restringe el área a las relaciones laborales entre gerencia y empleados. La tercera pregunta busca detectar cuál es la opinión del sujeto respecto de las relaciones laborales entre gerentes y empleados, y la quinta se torna más específica al preguntar si el sujeto considera que las relaciones deberían ser más cordiales o más distantes.

Maintz, Holm y Hübner (1992), por su parte, señalan que el embudo de preguntas permite “obligar” al encuestado a responder preguntas sobre temas que pueden resultarle desagradables. Una vez que el encuestado respondió la pregunta general quizás le resulte más fácil responder las preguntas específicas.

2.9.3.9 Preguntas de cantidad (escalares)

¿Para qué proporción de sus actividades como comunicador en la empresa recibe usted instrucciones claras y específicas por parte de sus superiores? (Pregunta adaptada de Lininger y Warwick, 1978, 193)

Recibo instrucciones claras y específicas

- Para casi todas mis actividades
- Para la mayor parte de mis actividades
- Para la mitad
- Para muy pocas de mis actividades
- Para casi ninguna de mis actividades

2.9.3.10 Preguntas de intensidad (o escala de Lickert)

Otro ejemplo es la escala de Lickert. La escala de Lickert presenta una afirmación o “declaración de actitudes” (Lininger y Warwick, 1978, 193) y el encuestado debe seleccionar una opción dentro de una secuencia continua que va desde “estoy completamente de acuerdo” hasta “estoy en total desacuerdo”.

La agenda mediática de XY cadena televisiva omite información proveniente de Cuba.

- Estoy completamente de acuerdo
- Estoy de acuerdo
- No estoy seguro
- No estoy de acuerdo
- Estoy en total desacuerdo

Otro ejemplo:

¿Cómo considera usted la atención recibida en el servicio telefónico de atención de clientes?

- Muy buena
- Buena
- Regular
- Mala
- Muy mala

Otro ejemplo de pregunta medida según la Escala de Lickert:

Por favor, señale su grado de acuerdo con las siguientes afirmaciones. Marque con una cruz el casillero que mejor refleje su opinión. (Pregunta adaptada de Hayes, 1998,161).

	Muy de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Muy en desacuerdo	No contesta
Ante todo estoy satisfecho con la calidad de los artículos publicados por XX						
Los artículos en XX contienen opiniones de los periodistas que desmerecen en parte la calidad de los mismos						
Considero que la información brindada en los artículos es muy confiable						
Me agrada la mirada crítica que mantienen los periodistas en sus artículos						
Los artículos son tendenciosos						

Algunas categorías alternativas para armar un escalamiento de Lickert son las que sugiere Hernández Sampieri, Fernández Collado y Baptista Lucio (1998, 257):

Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo
Definitivamente sí	Probablemente sí	Indeciso	Probablemente no	Definitivamente no
Completamente verdadero	Verdadero	Ni falso, ni verdadero	Falso	Completamente falso

2.9.3.11 Preguntas de frecuencia

Similares a las preguntas de cantidad, éstas comienzan por “con qué frecuencia...”

¿Con qué frecuencia lee usted la revista *Noticias*?

- mensualmente
 cada 2 meses
 cada 1 mes y medio
 no contesta

2.10 Formas de administración del cuestionario

Una vez construido el cuestionario, hay que aplicarlo. Existen distintas formas de aplicación (Sampieri, Collado y Lucio, 2000, 291). Veamos algunas de ellas:

1. Cuestionario auto-administrado (291): Se lo proporciona directamente en mano a los encuestados y ellos deben responderlo por cuenta propia, sin intermediarios, siguiendo las indicaciones correspondientes. Las indicaciones deberán ser claras para que el encuestado sepa cómo responder las preguntas. El cuestionario autoadministrado no podrá aplicarse ni a personas analfabetas, ni a niños ni a personas

con dificultades lectoras en general.

2. Entrevista personal (291): El cuestionario es aplicado por un entrevistador. El encuestador debe formular las preguntas de manera oral a los sujetos que estudia y debe anotar en la cédula las respuestas que obtiene. Aquí las instrucciones sobre cómo llenar el formulario son para el encuestador. Éste deberá tener cabal conocimiento sobre cómo proceder con el instrumento de medición. Es recomendable que el sujeto que realice la encuesta haya recibido con antelación entrenamiento en el arte de entrevistar. El encuestador deberá esforzarse por: lograr aceptación por parte de los encuestados; generar empatía; evitar inducir respuestas; evitar incomodar a los sujetos encuestados.
3. Por entrevista telefónica (291): El cuestionario se aplica vía telefónica. También en este caso es necesario un entrenamiento previo por parte del encuestador. La entrevista telefónica no es altamente recomendable puesto que consiste en trabajar con muestras autoelegidas, o lo que es igual, con muestras no probabilísticas seleccionadas según el criterio y conveniencia del investigador. En este sentido, los resultados obtenidos son sesgados. Se recomienda al lector remitirse a la sección 1.7.2 Marco de Muestreo para afianzar esta noción.
4. Auto-administrado y enviado por correo postal, electrónico o servicio de mensajería (291): Los encuestados reciben el cuestionario por correo electrónico, por correo postal o por otro medio y deben responderlo por sí mismos sin intermediarios. Al igual que en el caso de los cuestionarios auto administrados, las indicaciones de llenado del formulario deben estar bien redactadas, ser claras y precisas, puesto que el investigador no estará presente en caso de dudas por parte del sujeto encuestado.

5. Por Internet: El cuestionario, localizado en una página Web, es respondido directamente por el encuestado *on line*.

2.11 Procesamiento de la información: numeración de cédulas y codificación

Luego de la aplicación del cuestionario a la totalidad de los sujetos que conforman la muestra, debemos reunir todas las cédulas y procesar los datos que se encuentran en ellas.

Este procedimiento requiere de varias etapas:

- a) Numerar todos los cuestionarios desde 1 hasta n (siendo n el total muestral). Para numerar los cuestionarios, podemos, previamente, ordenarlos alfabéticamente según sea el apellido de los sujetos encuestados (si conocemos su nombre y apellido). Por ejemplo, al cuestionario correspondiente al sujeto de apellido Acevedo, le colocamos el número 1. Al del sujeto de apellido Adriano, le colocamos el número 2, y así sucesivamente, hasta llegar al último cuestionario que se corresponderá con el sujeto cuyo apellido empiece con una letra del abecedario más avanzada. Cuando no sabemos los apellidos de nuestros encuestados, también numeramos los cuestionarios.
- b) Codificar las respuestas de los cuestionarios. Este procedimiento puede haberse formalizado previamente en la etapa de construcción del instrumento de medición. Sin embargo, en ocasiones, la codificación se realiza en la etapa de análisis de datos. Codificar es, según Hernández Sampieri, Fernández Collado y Baptista Lucio (2000, 319), “transformar las

preguntas y respuestas de un cuestionario en símbolos o valores numéricos”. La codificación es un modo de preparar los datos para su posterior análisis estadístico, fundamentalmente, porque los programas estadísticos de tratamiento de datos trabajan con códigos y no con los nombres de las variables.

2.12 Codificación

La codificación de las preguntas y respuestas consiste en 4 pasos fundamentales:

2.12.1 Codificación de preguntas y categorías:

Este paso consiste en “codificar las categorías de ítems, preguntas y categorías de contenido u observación no pre codificadas” (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000, 319). Por ejemplo, supongamos que la décima pregunta de un cuestionario es: “Usted vive en...”, y las opciones de respuesta son: “casa; departamento; vivienda en lugar de trabajo; inquilinato; hotel o pensión; construcción no destinada a fines habitacionales; vivienda en villa”. Veamos la codificación:

10. Usted vive en
 - 1 Casa
 - 2 Departamento
 - 3 Vivienda en lugar de trabajo
 - 4 Inquilinato
 - 5 Hotel o pensión
 - 6 Construcción no destinada a fines habitacionales
 - 7 Vivienda en villa

2.13 Libro de códigos

El libro de códigos (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000, 320) es un documento que especifica la manera en que se ubicarán las variables y los códigos correspondientes a sus categorías en lo que será la matriz de datos. El mismo:

- a) Determina la manera de construir la matriz de datos en el programa estadístico;
- b) Establece los “significados de los valores de las categorías de las variables” (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000, 320).

Los elementos constitutivos de un manual de códigos son: variable (nombre de la variable o pregunta), categorías (dimensiones de la variable o, lo que es lo mismo, categorías de respuesta), códigos, números de columna, naturaleza de la variable (si es cualitativa o cuantitativa). A continuación un ejemplo de un fragmento de libro de códigos con la variable “Usted vive en”.

Variable	Categorías	Códigos	Columna	Naturaleza
10. Tipo de Vivienda	Casa	1	10	Variable Cualitativa
	Departamento	2		
	Vivienda en lugar de trabajo	3		
	Inquilinato	4		
	Hotel o pensión	5		
	Construcción no destinada a fines habitacionales	6		
	Vivienda en Villa	7		

Este segmento del manual de códigos nos sirve para saber que la variable “Tipo de vivienda”, es una variable cualitativa, que las opciones de respuesta son 7 y se ubican en la columna 10 de la

matriz de datos. Además, podemos saber el significado de cada valor de cada categoría de la variable: el valor 7 de la variable “vive usted” se corresponde con la categoría “vivienda en villa”.

2.14 Codificación física

A partir del manual de códigos podemos construir una matriz en un programa de tratamiento cuantitativo de datos como EXCEL, SPSS, INFOSTAT. Son documentos que contienen la totalidad de los datos recabados en el campo. En palabras de Hernández Sampieri, Fernández Collado y Baptista Lucio (2000, 321) la matriz de datos es “el conjunto de datos simbólicos o numéricos producto de la aplicación del instrumento”.

La matriz se construye siguiendo las indicaciones del manual de códigos e implica cargar los datos en un “documento permanente” (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2000). Veamos un segmento de matriz de datos en el cual hemos ubicado a la variable “Tipo de vivienda”:

Columna 1	Columna 6	Columna 7	Columna 8	Columna 9	Columna 10
Nombre	Variable 6	Variable 7	Variable 8	Variable 9	Tipo de vivienda
Acevedo, J.	2	1	5	2	4
Adriano, L.	4	1	1	2	1
Bilardo, J.	2	2	1	1	2
Francesca, D.	1	2	1	2	4

Esta tabla que aparentemente es incomprensible, muestra que, el señor Acevedo, vive en una casa o departamento alquilado (código 4 en la matriz, que se corresponde con la categoría “casa o departamento alquilado” de la variable “Vive Usted...”), en cambio, el señor Adriano en casa propia (código 1 en la matriz). A diferencia de estos dos señores, Bilardo vive en un departamento propio (código 2 en la matriz). Y así podemos seguir leyendo los datos. En definitiva,

los códigos numéricos representan las opciones de respuesta que los sujetos han marcado en cada uno de los cuestionarios. Si miramos el segmento de matriz horizontalmente, vemos que, en la primera fila tenemos toda la información correspondiente al señor Acevedo, y en la segunda toda la información correspondiente al señor Adriano, y así sucesivamente. Mientras tanto, si miramos la matriz verticalmente, encontramos que, por ejemplo, en la columna 10 tenemos toda la información referida a la variable número 10 (Tipo de vivienda).

2.15 Matriz de datos

Hernández Sampieri, Fernández Collado y Baptista Lucio (2000, 319 y 329) señalan que el cuarto paso consiste en grabar y guardar los datos de la matriz en un documento o archivo permanente. El archivo permanente es generado desde el momento en que abrimos un documento en el programa estadístico para construir la matriz. Resulta de vital importancia generar copias de seguridad de la matriz, no sólo en la misma máquina, sino también en CDs, DVs, Pen Drives y/o archivos electrónicos ubicados en algún sitio de internet (casilla de correos, entre otros).

2.16 Codificación manual

En caso que no se sepa utilizar programas estadísticos como INFOSTAT, EXCEL, entre otros, puede concebirse, para el procesamiento y posterior análisis de los datos, una “codificación manual” (Apezechea, 1979, 125-126). En este caso, se revisan las fichas de encuesta una por una y se van categorizando en una hoja aparte las respuestas obtenidas. Tomemos por ejemplo la variable “Estado civil” con seis opciones posibles de respuesta: soltero, casado, divorciado/separado, viudo, sin dato. Primero, tomamos una hoja y armamos una tabla que contenga las seis posibilidades de respuesta. Segundo, revisamos cada cuestionario o ficha y anotamos con un tilde la respuesta obtenida en la celda correspondiente (si

en la primera ficha el sujeto seleccionó la opción soltero, entonces hacemos un tilde en la celda que dice soltero y así sucesivamente). Por último, contabilizamos la cantidad de casos que caen en cada una de las categorías de respuesta y anotamos el total en la celda correspondiente en la columna “Cantidad”. Abajo un ejemplo ilustrativo:

Variable “Sexo”	Cantidad
soltero //////////////////////////////////////	20
casado //////////////////////////////////////	40
divorciado/separado //////////////////////////////////////	20
viudo //////////////////////////////////	10
sin dato //////////////////////////////////	10
TOTAL	100

Luego de realizar este procedimiento, que evita el proceso de codificación, se procede a trasladar los datos a una planilla de cálculo y se extraen los porcentajes. Habitualmente, cuando se trabaja de este modo no se construyen bases de datos, hecho que dificulta un tratamiento exhaustivo y complejo de los datos recabados en una encuesta.

Una vez que tenemos construida la matriz podemos entonces realizar el análisis estadístico de los datos.

3. RUDIMENTOS DE ESTADÍSTICA DESCRIPTIVA

3.1 Análisis y lectura de datos

El análisis estadístico de los datos depende de tres factores fundamentales. Por una parte, del nivel de medición de las variables; por la otra, de la manera como se hayan formulado las hipótesis; por último, del interés que posea el investigador.

La manera de realizar análisis estadísticos varía según ésta sea manual o a través de programas informáticos. En el presente manual se desarrollan todos los análisis a partir de los procedimientos manuales. Ahora bien, es fundamental aclarar que todo lo desarrollado en el presente capítulo puede realizarse a través de programas como INFOSAT; SPSS, STATA o EXCEL.

3.1.1 Tabla de frecuencias

Si trabajáramos desde un programa estadístico, le pediríamos al programa que extraiga tablas de frecuencias como primera medida para realizar un análisis descriptivo de las variables en estudio. En síntesis, lo que el investigador obtiene, en este sentido es la distribución de frecuencias obtenidas para cada variable en particular.

La información que puede aparecer en una tabla de frecuencias es básicamente la que presentamos a continuación:

Nombre de la variable y nombre de cada categoría o clase. (Ejemplo: variable edad)	Frecuencia absoluta de clase (o cantidad de observaciones por categoría de la variable)	Frecuencia relativa de clase (o frecuencia absoluta de cada categoría de la variable dividido el total de observaciones)	Porcentajes (o frecuencia relativa de cada categoría de la variable multiplicada por 100)	Porcentaje válido	Frecuencia absoluta acumulada de clase (o sumatoria de las frecuencias absolutas)	Frecuencia relativa acumulada de clase (o sumatoria de las frecuencias relativas)
15 a 19 años	3	$3/40 = 0,075$	$0,075 * 100 = 7,5\%$	7,5	3	$3/40 = 0,075$
20 a 24 años	7	$7/40 = 0,175$	$0,175 * 100 = 17,5\%$	17,5	$(3+7) = 10$	$10/40 = 0,250$
25 a 29 años	8	$8/40 = 0,200$	$0,200 * 100 = 20\%$	20	$(10+8) = 18$	$18/40 = 0,450$
30 a 34 años	8	$8/40 = 0,200$	$0,200 * 100 = 20\%$	20	$(18+8) = 26$	$26/40 = 0,650$
35 a 39 años	7	$7/40 = 0,175$	$0,175 * 100 = 17,5\%$	17,5	$(26+7) = 33$	$33/40 = 0,825$
40 a 44 años	7	$7/40 = 0,175$	$0,175 * 100 = 17,5\%$	17,5	$(33+7) = 40$	$40/40 = 1,000$
Total	40	1	100%		40	1

Y... ¿cómo leemos los datos de esta tabla? El 40% de los sujetos tienen entre 25 y 34 años, mientras que otro 35% entre 35 y 44 años. El restante 25% (o lo que es lo mismo, $\frac{1}{4}$ de los sujetos) entre 15 y 24 años. El grupo más numeroso de sujetos tiene edades que oscilan entre los 25 y 34 años y el menos numeroso entre 14 y 24 años.

Definamos ahora cada uno de los conceptos mencionados en el cuadro. La frecuencia absoluta de una variable es “el número de veces que se repite cada categoría de la variable” (Blanch y Joekes, 1997, 17). Mientras tanto, entendemos por *frecuencia relativa* a la “proporción de veces que ocurre cada categoría de la variable”, o lo que es igual, el número de veces que se repite la categoría de la variable “dividido por el número total de observaciones” (Blanch y Joekes, 1997, 21). Por último, la frecuencia acumulada es “la suma de las frecuencias absolutas o relativas de los valores menores o iguales al valor que se está considerando” (Blanch y Joekes, 1997, 21).

Generalmente, en las tablas que arrojan los programas estadísticos no encontramos columnas en las que aparezca la frecuencia relativa. Directamente hallamos las columnas de

porcentajes y porcentajes válidos. La columna de porcentaje válido es de vital importancia para conocer en qué medida podemos o no leer los datos de la tabla. Cuando el porcentaje válido es diferente al porcentaje, esto indica que hay en la base de datos casos perdidos (o *missing cases*) y que debemos guiarnos por la columna de porcentaje válido en vez de por la columna porcentaje.

La tabla presentada se denomina tabla de frecuencias. Una tabla de frecuencias, en definitiva, “asocia a cada categoría o valor de la variable con las veces que se repite dicha categoría o valor” (Blanch y Joekes, 1997, 15 y 20).

A la información presente en una tabla podemos graficarla. Cuando trabajamos con variables cualitativas categóricas, podemos realizar un gráfico de barras (*bar chart*) o de tortas (*pie chart*). Cuando trabajamos con variables cuantitativas discretas también podemos realizar un gráfico de barras (*bar chart*), mas no un gráfico de tortas. Por el contrario, cuando estamos ante variables cuantitativas continuas, optamos por un histograma (*histogram*).

3.1.1.1 Tablas de frecuencias para variables categóricas dicotómicas

Las tablas más simples para el análisis son las de variables categóricas dicotómicas. Recordemos que una variable dicotómica tiene sólo dos categorías, una es sí y la otra es no. Medimos la presencia o ausencia en las unidades de análisis de determinada característica cualitativa. Imaginemos que, queremos averiguar si nuestros clientes comprarían en el futuro algún otro producto de la empresa y le preguntamos: “Ud, ¿compraría más adelante algún otro producto de esta empresa?” Las respuestas de los entrevistados serán, lógicamente, sí o no. Veamos cómo quedaría sistematizada la información en una tabla:

1. ¿Compraría más adelante un producto de ésta empresa?

Columna 1	Cantidad	Porcentaje
Sí	56	39%
No	88	61%
Total	144	100%

¿Cómo leemos los datos? Es muy simple. En este caso, de un total de 144 clientes, más de la mitad (un 61%) señaló que no compraría productos de la empresa en un futuro; dato altamente significativo, si se quiere mejorar la calidad del producto, su imagen en el mercado y consecuentemente, su nivel de ventas.

3.1.1.2 Tablas de frecuencias para variables cualitativas

Veamos el caso de una variable categórica con más de dos opciones de respuesta, como *estado civil*:

2. Estado civil de los alumnos de la carrera de Comunicación Social

Estado Civil	Cantidad	Porcentaje	Porcentaje Válido
Soltero	25	25%	25%
Unido	30	30%	30%
Casado	25	25%	25%
Separado/ divorciado	15	15%	15%
Viudo	5	5%	5%
Total	100	100%	100%

En relación con esta tabla, es interesante subrayar que hay mayor cantidad de sujetos unidos de hecho que casados y solteros. Sin embargo, la diferencia porcentual entre los sujetos unidos de hecho y los que están solteros o están casados no es significativa. Mientras que los unidos de hecho llegan a un 30%, los casados a un 25% y los solteros a otro 25%. Podemos decir entonces que, el 50% de los sujetos son casados o solteros, y que un 30% son unidos de

hecho. Otra manera de leer los datos podría ser la siguiente: mientras la mayoría (un 55% del total de 100 sujetos) es unido de hecho o casado, $\frac{1}{4}$ es soltero. El porcentaje de personas separadas es, en cambio, de un 15%; porcentaje levemente menor.

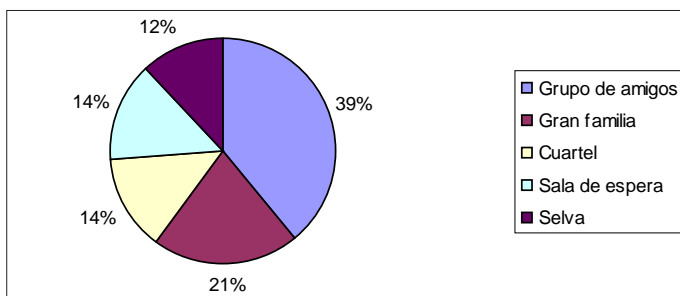
3.1.1.3 Tablas de frecuencia para variables cualitativas

Otro aspecto que puede estudiarse es cómo perciben los empleados el clima interno de la empresa. Imaginemos que en el caso referido se planteó en cuestionario una pregunta sobre cómo percibía el empleado el clima de trabajo en la empresa y que las opciones de respuesta (mutuamente excluyentes) fueron las siguientes: “es como un grupo de amigos; es como una gran familia; es como un cuartel; es como una sala de espera; es como una selva”. Analicemos la siguiente tabla:

Percepción del clima interno de la empresa por parte de los empleados

Clima interno	Cantidad	Porcentaje	Porcentaje Válido
Grupo de amigos	39	39%	39%
Gran familia	21	21%	21%
Cuartel	14	14%	14%
Sala de espera	14	14%	14%
Selva	12	12%	12%
	100	100%	100%

Evaluación del Clima laboral de la empresa



Más de la mitad de los sujetos (un 60%) definió al clima del ámbito de trabajo de manera positiva: como similar al de un grupo de amigos o al de una gran familia. Mientras tanto, un 40% de los encuestados señaló que el clima interno de la empresa se asemeja al de un cuartel, al de una sala de espera o al de una selva. Interesante es ver aquí el contraste significativo entre la manera de entender al clima de la empresa por parte del grupo mayoritario y el segundo grupo. Quizás un investigador inquieto se propondría, a partir de estos resultados, estudiar más en profundidad al 40% de sujetos que considera al clima laboral como un cuartel, una sala de espera, o una selva. Inclusive, esto puede además estudiarse con metodologías cualitativas.

La interpretación de los datos estadísticos dependerá siempre de los objetivos de investigación que nos planteamos al inicio del trabajo. Es decir, si el objetivo de una encuesta es obtener datos que nos permitan formular planes de acción para mejorar una situación o una realidad, seguramente tendremos que priorizar ciertos datos por sobre otros (quizás debamos resaltar aquéllos que no resultan del todo agradables, pero que muestran con claridad ciertos aspectos que no pueden ser soslayados).

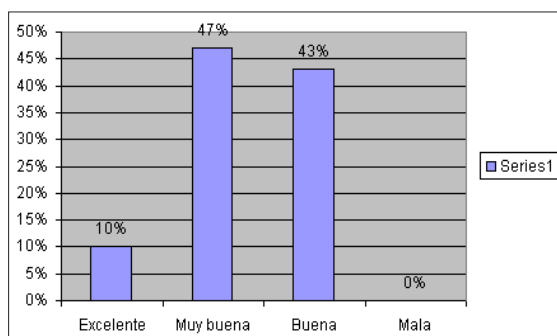
3.1.1.4 Tablas de frecuencia para variables cualitativas ordinales

Supongamos ahora que hicimos una encuesta para conocer la percepción que los empleados de una empresa tienen respecto de sus posibilidades de progresar en el trabajo y que los resultados son los que se presentan a continuación:

1. Perspectiva de progreso de los empleados en la empresa

Perspectiva de Progreso	Cantidad	Porcentaje	Porcentaje Válido
Excelente	10	10%	10%
Muy buena	47	47%	47%
Buena	43	43%	43%
Mala	0	0%	0%
Total	100	100%	100%

Perspectivas de progreso en la empresa por parte de los empleados



Una manera de leer el gráfico podría ser: un 57% de sujetos señaló que su perspectiva de progreso en la empresa es entre excelente y muy buena, mientras que, un 43% señaló tener buenas perspectivas. La diferencia porcentual entre estos dos grupos no es muy significativa. Resulta en algo muy positivo el hecho de que no haya personas que consideren que no tienen posibilidades de progresar.

Otra manera de leer el gráfico es contrastando las dos categorías modales. Mientras un 47% de los sujetos sostiene que tiene muy buenas perspectivas de progreso; otro 43% sostiene que esas perspectivas son buenas.

3.1.1.5 Tablas de contingencia o bidimensionales

La lectura de las tablas y gráficos anteriores resulta relativamente sencilla. Pasemos ahora a la lectura de tablas de doble entrada, las cuales suele traer aparejadas ciertas dificultades para su total comprensión.

Al programa de tratamiento estadístico de datos podemos también pedirle que arroje tablas de doble entrada en las cuales se cruzan los datos correspondientes a dos variables distintas, o incluso tres. Observemos la siguiente tabla:

4. Máximo nivel educativo de clientes mujeres y varones (en números)

Máximo Nivel Educativo	Mujeres	Varones	Total
Universitario	58	59	117
Maestría	10	10	20
Doctorado	28	26	54
Otros	4	5	9
Total	100	100	200

A partir de esta tabla, en la que tenemos sólo las frecuencias absolutas para cada categoría de las variables, hay dos posibilidades de lectura: por un lado, la lectura horizontal y por el otro la lectura vertical. Podemos decir, ciertamente, que de 117 clientes con título universitario, 58 son mujeres y 59 son varones; que de 20 clientes con estudios de maestría, 10 son mujeres y 10 son varones, y así sucesivamente. También podemos decir que, de 100 clientas mujeres, 58 tienen título universitario, 28 título de doctorado, 10 de maestría y sólo 4 tienen otro tipo de títulos. El mismo tipo de lectura podemos hacer con los datos sobre los clientes varones.

La extracción de porcentajes puede realizarse tanto de manera vertical como horizontal. En este caso veamos cómo extrajimos los porcentajes teniendo en cuenta la variable “sexo”.

5. Máximo nivel educativo de clientes mujeres y varones (en porcentajes)

Máximo Nivel Educativo	Mujeres	Varones
Universitario	58%	59%
Maestría	10%	10%
Doctorado	28%	26%
Otros	4%	5%
Total	100%	100%

Aquí, la lectura puede realizarse del modo siguiente: la mayoría de los clientes mujeres y varones tiene título universitario (un 58% para ellas y un 59% para ellos). Algo más de $\frac{1}{4}$ de los varones (un 26%) y un 28% de las mujeres tiene título de doctorado. Coincidentemente, tanto un 10% de mujeres como de varones tienen título de maestría. Como conclusión, podemos destacar que es llamativo el alto porcentaje de clientes mujeres y varones que han finalizado estudios de grado y postgrado.

3.1.2 Medidas de tendencia central

A los conjuntos de datos podemos describirlos, además, en función de medidas de tendencia central. La tendencia central de cualquier conjunto de datos es la disposición de éstos para agruparse ya sea alrededor del centro o de ciertos valores numéricos (Blanch y Joekes, 1997c). Existen, principalmente, tres medidas de tendencia central: la moda, la mediana y la media.

3.1.2.1 La moda

La moda es una medida de posición, es decir, posiciona a un conjunto de datos. Particularmente, la moda es “aquel valor o aquella categoría de la variable que ocurre o se presenta con la mayor frecuencia (Blanch y Joekes, 1997c, 4). La moda es una categoría de la variable y no la frecuencia de aparición de esta categoría.

Supongamos que estamos analizando la variable cualitativa “Máximo nivel educativo alcanzado” (Blanch y Joekes, 1997c, 4) por parte de los clientes de una empresa y hallamos que la categoría “Nivel primario completo” es la de mayor frecuencia de aparición; entonces decimos que la categoría modal de la variable “Máximo nivel educativo” es “Nivel primario completo”, y no que la moda es por ejemplo, el número total de observaciones que concentra esa categoría. Otra manera de leer esta medida de tendencia central es diciendo que la categoría con mayor frecuencia de aparición de la variable “Máximo nivel educativo” para los clientes de la empresa

es “Nivel primario completo”. O, también podemos plantearlo del siguiente modo: “el grupo más numeroso de clientes de la empresa xx tiene estudios primarios completos/ o ha completado sus estudios primarios”. Si al inicio de la investigación el investigador esperaba encontrar, por ejemplo, un mayor porcentaje de clientes con estudios secundarios completos en vez de con estudios primarios completos, entonces uno podría leer el dato del siguiente modo: “el grupo más numeroso de clientes de la empresa sólo posee estudios primarios completos”.

3.1.2.1.1 Moda para variables cuantitativas discretas

Veamos ahora cómo leer la moda para variables cuantitativas discretas. Supongamos que hemos decidido encuestar a un total de 44 individuos de los cuales, 20 son mujeres y 24 varones. Imaginemos también que en el cuestionario aplicado, además de la pregunta sobre si el sujeto es mujer o varón, colocamos otra pregunta relativa a la cantidad de ponencias presentadas en congresos internacionales el cabo de un año. Luego de armar la base de datos, le pedimos al programa estadístico que arroje una tabla de contingencia para analizar las categorías modales. Veamos la tabla (tomada y adaptada de Blanch y Joekes, 1997c):

2. Cantidad de ponencias presentadas en congresos internacionales según sexo

Número de ponencias	Mujeres	Varones	Total
0	12	5	17
1	5	15	20
2	1	3	4
3	2	1	3
4	0	0	0
Total	20	24	44

En este caso tenemos que, la categoría modal para las mujeres es de 0 ponencias, o lo que es lo mismo, el grupo más numeroso de mujeres no presentó ponencias en congresos internacionales. En relación con los varones, en cambio, el grupo más numeroso presentó 1 ponencia en todo el año.

3.1.2.1.2 Moda para variables cuantitativas continuas

Cuando trabajamos con variables cuantitativas continuas, tales como nivel de ingresos mensuales, los datos suelen agruparse en intervalos de clase y la moda se determina en función de la frecuencia de aparición de dichos intervalos. La moda es, en este caso, el intervalo de clase con mayor frecuencia de aparición.

Supongamos que estamos analizando la variable “promedio de ingresos mensuales” de los clientes de una empresa y tenemos la siguiente distribución (Blanch y Joeques, 1997c):

6. Promedio de ingresos mensuales de clientes

Nivel promedio de ingresos mensuales	Frecuencia absoluta (ó nj)
De 400 a 800	5(nj-1)
De 801 a 1200 (Intervalo modal)	20 (nj)
De 1201 a 1600	17 (nj+1)
De 1601 a 2000	10
De 2001 a 2400	9
De 2401 a 2800	8
Más de 2801	7

El cálculo de la moda puede realizarse siguiendo dos procedimientos diferentes. Uno de ellos implica tomar “el punto medio del intervalo modal” (Blanch y Joeques, 1997c,10) como la moda. En el ejemplo de la tabla, el intervalo modal sería “De 801 a 1200” (es la categoría de la variable con mayor frecuencia de aparición), y la moda 1000 (es el punto medio entre 800 y 1200).

Esto significa que el grupo más numeroso de clientes de la empresa x gana aproximadamente unos 1000 pesos mensuales.

El otro cálculo es algo más sofisticado. En este caso tenemos que realizar operaciones matemáticas. La moda aquí es igual a: $Y_{j-1} + c * (f_1 / f_1 + f_2)$, donde Y_{j-1} representa al extremo inferior del intervalo modal; c a la amplitud del intervalo modal; f_1 a la frecuencia del intervalo modal menos la frecuencia del intervalo anterior, es decir, representa a $(n_j - n_{j-1})$; f_2 a la frecuencia del intervalo modal menos la frecuencia del intervalo posterior, o lo que es lo mismo, representa a $(n_j - n_{j+1})$.

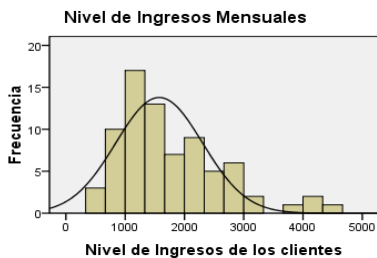
Para el ejemplo de la tabla, tenemos los siguientes números:

- $Y_{j-1} = 801$ (o lo que es igual, Y_{j-1} es el extremo inferior del intervalo modal. En este caso, el intervalo modal es “de 801 a 1200”, por lo tanto el extremo inferior del intervalo es 801)
- $C = 400$ (o lo que es lo mismo, c es igual a la amplitud del intervalo, que se extrae restando el extremo superior menos el extremo inferior del primer intervalo, en este caso, $800 - 400 = 400$)
- $f_1 = n_j - n_{j-1} = 20 - 5 = 15$ (Frecuencia del intervalo modal -en este caso el intervalo modal es “de 801 a 1200” y la frecuencia absoluta de 20- menos la frecuencia del intervalo anterior -el intervalo anterior es “de 400 a 800”, y la frecuencia de 5-)
- $f_2 = n_j - n_{j+1} = 20 - 17 = 3$ (Frecuencia del intervalo modal menos la frecuencia del intervalo posterior)

La fórmula dijimos que era: $Y_{j-1} + c * (f_1 / f_1 + f_2)$. Tras el reemplazo correspondiente, tenemos: moda = $801 + 400 * (15 / 15 + 3) =$

A partir de esta operación matemática podemos decir que el grupo más numeroso de clientes de la empresa gana alrededor de \$1134 mensuales.

El gráfico correspondiente a la variable “Nivel de ingresos” promedio de los clientes es un histograma. En él, la primera barra que aparece representa a la primera categoría de la variable “de 400 a 800”, y la última a la última categoría de la variable “más de 2800”.



Se utiliza un histograma porque refleja la condición de continuidad de la variable cuantitativa. Es decir, la variable se subdivide en intervalos que se encuentran ligados unos a otros. Donde termina el primer intervalo empieza el que sigue. Por eso las barras están unidas y no separadas como en los gráficos de barras donde cada categoría de la variable no tiene relación con la que sigue (puesto que en esos casos trabajamos con variables categóricas o variables cuantitativas discretas).

3.1.2.1.3 Ejemplos sobre cómo trabajar datos categóricos analizando los grupos modales

Resultados del análisis de las evaluaciones de la VI Bienal Iberoamericana de Comunicación, ECI, UNC, año 2007.

Responsable: Cecilia Blanco

En primer lugar, aparentemente, el desarrollo de la Bienal ha sido considerado como positivo por gran parte de los asistentes.

En relación con la pregunta acerca de la opinión que le mereció al participante el desarrollo de la VI Bienal, encontramos que la categoría de respuesta con mayor frecuencia de aparición es la de “muy buena”. Coincidentemente, y en relación con la manera de evaluar la calidad de los temas propuestos en las mesas de trabajo hallamos que la mayoría de los sujetos indicó también que éstos eran “muy buenos”. Lo mismo sucede con la pregunta sobre la calidad de los temas propuestos en las conferencias: la categoría elegida con mayor frecuencia fue “muy buenos”.

La mayor parte de los encuestados indicó que los horarios de las mesas de trabajo se respetaron “casi siempre”; y que los temas propuestos por los disertantes fueron “casi siempre” claros.

Según los datos arrojados por la tabla N° 1, la mayor parte de los participantes que devolvieron el cuestionario fueron mujeres. El grupo mayoritario de estos participantes tiene hasta 25 años. La mayoría es de Córdoba Capital y particularmente representan a la institución Escuela de Ciencias de la Información, de la Universidad Nacional de Córdoba.

Extracto del artículo “El perfil social y económico de los estudiantes-ingresantes a cuatro carreras de la Facultad de Filosofía y Humanidades de la UNC del año 2004”, publicado en Actas de las V Jornadas de Investigación en Educación, julio de 2007. Autora: Cecilia Blanco.

En primer lugar, intentaré describir a los estudiantes según las variables “Tipo de estudios secundarios realizados” y si tienen o no otro tipo de estudios superiores previos a la carrera en la que se matricularon. Veamos a continuación los resultados del análisis.

En relación con el año de egreso del secundario podemos afirmar que la mayoría de los ingresantes egresó del secundario entre los años 1994 y 2003. Para Historia, el porcentaje de egresados en estos años es de un 81%; para Letras de casi un 78%, para Filosofía de un 72% y para Ciencias de un 61%. Como puede observarse el

porcentaje de egresados entre 1994 y 2003 es significativamente más bajo en Ciencias que en las demás carreras. Mientras tanto, en Ciencias, hay un segundo grupo numeroso de alumnos que egresó del secundario entre 1984 y 1993: un 23%, mientras que en las demás carreras hallamos porcentajes que oscilan entre el 6% en Filosofía y el 8% en Historia. Otro grupo relativamente importante de alumnos de Ciencias egresó entre 1974 y 1983: un 10%, mientras que en Filosofía un 5%, en Letras un 4% y en Historia apenas un 3%.

En cuanto al tipo de estudios secundarios cursados por estos ingresantes, podemos señalar que alrededor de $\frac{1}{4}$ de ellos indicó haber realizado estudios secundarios de tipo "Bachiller Humanista": un 33% de Filosofía; un 28% de Ciencias; un 24% de Historia y un 19% de Letras. Ahora bien, porcentajes cercanos de alumnos indicaron haber cursado estudios secundarios de tipo "Bachiller Comercial": un 23% de Ciencias; un 20% de Historia, un 13% de Letras y un 10% de Filosofía. Por otra parte, mientras el tercer porcentaje importante para Letras fue el de los estudiantes que cursaron estudios secundarios de tipo de "Humanidades: Ciencias Sociales" con un 7%, para Filosofía e Historia fue el de los alumnos con estudios secundarios de tipo "Economía y Gestión de las Organizaciones: Gestión Administrativa", con un 7% para la primera y un 9% para la segunda. En cambio, para Ciencias, esta tercera categoría modal se corresponde con el grupo de alumnos que cursó un "Bachillerato Pedagógico", con un 9%.

Además de lo anteriormente señalado, resulta interesante observar el comportamiento de estos estudiantes antes de iniciar las carreras de Filosofía, Ciencias, Historia y Letras. En tres de las carreras, en Filosofía, en Letras y en Historia, encontramos que la mitad de sus ingresantes no tiene otros estudios superiores, un 45% para la primera, un 48% para la segunda y un 51% para la tercera; mientras que en Ciencias, el grupo más numeroso de alumnos finalizó estudios terciarios antes de ingresar a la carrera: un 35%. Por otra parte, si bien dijimos que en Filosofía, Letras e Historia la mayoría de los alumnos no cursa estudios superiores,

hallamos en estas carreras un segundo grupo numeroso que sí continúa estudios universitarios: un 18% para la primera, un 20% para la segunda y un 16% para la tercera. Mientras tanto, en Ciencias este segundo grupo importante señaló no tener otros estudios superiores. Asimismo, en Ciencias, el tercer porcentaje en importancia es el de los estudiantes que continúan estudios terciarios, de un 11%, y para las otras tres carreras, es el de aquellos ingresantes que abandonaron estudios universitarios: un 11% para Filosofía; un 10% para Historia, y un 6% para Letras.

3.1.2.2 La media aritmética

Estudemos ahora la media aritmética. La media aritmética de un conjunto de datos es una medida de posición denominada promedio (Blanch y Joeke, 1997c, 25). La media aritmética es la suma de todos los valores que toma una variable en estudio dividida por el número total de unidades de análisis.

3.1.2.2.1 Media para variables cuantitativas discretas (datos agrupados)

La fórmula que se utiliza para calcular la media aritmética de variables cuantitativas discretas agrupadas en una tabla es:

$$\text{media} = \sum (x_i * n_i) / n$$

Dónde:

- Σ significa sumatoria de;
- x_i representa los distintos valores que toma la variable;
- n_i es igual a la frecuencia absoluta correspondiente de cada valor de la variable
- n , es el número total de observaciones.

Ejemplo:

En una encuesta sobre satisfacción de los empleados se preguntó cuántas promociones recibió el sujeto durante el tiempo que trabajó en la empresa. Los resultados fueron los siguientes (Blanch y Joekes, 1997c):

7. Promociones dentro de la empresa

Promociones (xi)	Frecuencia absoluta o cantidad (ni)	(xi)(ni)
0	44	(0)(44)= 0
1	22	(1)(22)= 22
2	16	(2)(16) = 32
3	10	(3)(10) = 30
4	8	(4)(8) = 32
Total	100	116

$$\text{media} = \sum (x_i * n_i) / n$$

- Media = $\sum [(0)(44) + (1)(22) + (2)(16) + (3)(10)+(4)(8)] / 100 =$
- Media = $\sum [0 + 22 + 32 + 30 + 32] / 100 =$
- Media = $\frac{116}{100} =$
- Media = 1,16

La sumatoria de cada valor de la variable multiplicado por su frecuencia de aparición da igual, en este caso a 116. Si dividimos ese total por n (el total de datos), o sea, por 100, esto nos da: 1,16, lo cual quiere decir que en promedio los empleados de la empresa x han obtenido 1 promoción. Recordemos que el valor 1,16 es el resultado de un cálculo matemático y no un valor de la variable. No sería correcto plantear que en promedio los empleados tuvieron casi una promoción y media. Se trata de una variable cuantitativa discreta.

3.1.2.2.2 Media para variables cuantitativas continuas (datos agrupados)

Para el caso de variables cuantitativas continuas, el cálculo de la media es similar al anterior aunque con ciertas variaciones en la fórmula. En este caso particular, la fórmula es:

$$y = \sum (y_i * n_i) / n$$

Dónde:

- y_i es el punto medio del intervalo de clase
- n_i es la frecuencia absoluta de aparición de cada intervalo de clase
- n es igual al total de observaciones.

Ejemplo:

Imaginemos que, en la misma encuesta sobre satisfacción de los empleados suministrada en una empresa de 11 años de existencia, agregamos una pregunta como esta: “¿Cuánto tiempo lleva usted trabajando en esta firma?”, y que los resultados son los presentados en la tabla siguiente (Blanch y Joeques, 1997c):

8. Antigüedad de los empleados en la empresa

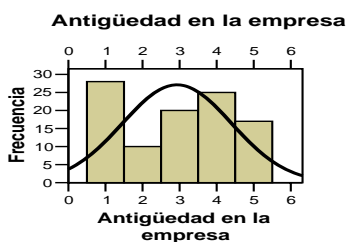
Cantidad de tiempo	Punto medio (y_i)	Frecuencia absoluta del intervalo de clase (n_i)	(y_i)(n_i)
Entre 1 y 3 años	2	28	(2)(28) = 56
Entre 3 y 5 años	4	10	(4)(10) = 40
Entre 5 y 7 años	6	20	(6)(20) = 120
Entre 7 y 9 años	8	25	(8)(25) = 200
Entre 9 y 11 años	10	17	(10)(17) = 170
Total		100	586

Ahora hagamos el cálculo de la media aritmética. Dijimos que la media aritmética de una variable cuantitativa continua es:

- $y = \Sigma (y_i * n_j) / n$
- $y = \Sigma [(2)(28) + (4)(10) + (6)(20) + (8)(25) + (10)(17)] / 100 =$
- $y = \Sigma [56 + 40 + 120 + 200 + 170] / 100 =$
- $y = \frac{586}{100} =$
- $y = 5,86$

Este resultado indica que en promedio los empleados de la empresa tienen una antigüedad de entre 5 y 6 años. Decimos entre 5 y 6 puesto que el resultado así lo indica. El 5,86 está más cercano a 6 que a 5, y como se trata de una variable cuantitativa continua entonces puede haber una serie de puntos intermedios entre el 5 y el 6: un empleado puede tener una antigüedad de 5 años y medio; otro de 5 años y 10 meses; etc.

Veamos ahora el histograma que grafica la información de la variable estudiada:



3.1.2.3 La mediana

La mediana es “un valor de la variable que divide a un conjunto de observaciones (ordenadas de mayor a menor) en dos subconjuntos que contienen la misma cantidad de datos” (Blanch y Joeques, 1997c, 13). Así, la mitad de las observaciones son menores o iguales que la mediana y la otra mitad mayores o iguales. La mediana no puede calcularse cuando trabajamos con variables categóricas medidas en escala nominal.

3.1.2.3.1 Mediana para variables cuantitativas discretas (datos agrupados)

Para variables cuantitativas discretas, las reglas para obtener la mediana son las siguientes:

- a) Cuando le pedimos al programa estadístico que arroje la tabla, le pedimos también, si no lo hace de manera automática, que extraiga “las frecuencias acumuladas correspondientes a cada valor de la variable (N_j)” (Blanch y Joeques, 1997c, 14), y, acto seguido, calculamos el orden de localización de la mediana estableciendo “el cociente $n/2$, donde n es el total de datos” (14).
- b) “Dentro de la columna correspondiente a las frecuencias acumuladas se busca la mayor frecuencia que no supera a $n/2$ y se simboliza con N_{j-1} ” (Blanch y Joeques, 1997c, 14). La frecuencia que le seguirá en orden será N_j . Si N_{j-1} es menor a $n/2$, entonces la mediana será igual a y_j , o sea el valor de la variable que posea frecuencia acumulada N_j . Por el contrario, si N_{j-1} es igual a $n/2$, entonces, la mediana será igual a $(y_{j-1} + y_j) / 2$, siendo y_{j-1} el valor de la variable con frecuencia N_{j-1} e y_j el valor de la variable con frecuencia N_j .

Para comprender lo anterior, volvamos a la tabla en la que analizamos la mediana para variables cuantitativas discretas, pero con algunas modificaciones.

8. Cantidad de revistas compradas en un mes por mujeres y varones

Número de Revistas compradas en un mes	Mujeres	N_j (Frecuencia acumulada)	Varones	N_j (Frecuencia acumulada) ²
0	3	3 (N_{j-1})	2	2 (N_{j-1})
1	17	20 (N_j)	16	18 (N_j)

2	1	21	5	23
3	2	23	1	24
4	0	23	0	24
Total	23		24	

Para las mujeres, $n/2$, es en números, $23/2= 11,5$. Buscamos ahora la mayor frecuencia acumulada que no supera a $n/2$ en el grupo de mujeres y hallamos la frecuencia 3. Le colocamos al lado N_{j-1} . A la frecuencia acumulada siguiente le colocamos N_j . La mediana será, en este caso, el valor de la variable que posee frecuencia acumulada N_j , es decir que la cantidad mediana de revistas compradas al mes es de 1 para las mujeres.

Para los varones $n/2$ es igual a $24/2= 12$. Buscamos entonces, la mayor frecuencia acumulada que no supere a 12. En este caso, es el 2. Colocamos, en primer término, N_{j-1} al lado de la frecuencia acumulada y en segundo lugar, N_j a la frecuencia siguiente. La categoría de la variable que coincide con la frecuencia acumulada N_j es la mediana. En este caso $Me=1$. Es decir, que la cantidad mediana de revistas compradas en un mes es de 1 también para los hombres. La mitad (o lo que es lo mismo, el 50%) compra una o menos de una revista al mes y la otra mitad, el otro 50% restante, compra una o más de una revista en un mes.

3.1.2.3.2 Mediana para variables cuantitativas continuas (datos agrupados)

Veamos ahora cómo tomamos la mediana para variables continuas. En este caso, también tenemos que resolver la siguiente fórmula matemática:

- $Me= Y_{j-1} + c [(n/2 - N_{j-1}) / (N_j - N_{j-1})]$,

Donde:

- $n/2$ = Número total de observaciones dividido 2
- N_{j-1} = Mayor de las frecuencias acumuladas que no supera a $n/2$
- N_j = Frecuencia acumulada que supera a $n/2$ y que le sigue a N_{j-1}
- Y_{j-1} = Extremo inferior del intervalo que tiene como frecuencia N_j
- c = Amplitud del intervalo (que se extrae restando el extremo superior menos el extremo inferior del primer intervalo)

Una de las variables cuantitativas continuas que podemos analizar es “Edad de los clientes”. Observemos la tabla (Blanch y Joeques, 1997c) que se presenta a continuación y hagamos los cálculos necesarios para obtener la mediana:

10. Edad de los clientes

Edad	Frecuencia relativa o cantidad (nj)	Frecuencia acumulada (Nj)
Entre 18 y 22	5	5
Entre 23 y 27	10	15
Entre 28 y 32	19	34
Entre 33 y 37	15	49 N_{j-1}
Entre 38 y 42	25	74 N_j
Entre 43 y 47	12	86
48 o más	14	100
Total	100	

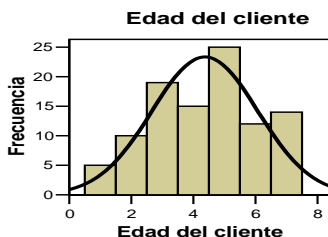
Calculemos ahora la mediana:

- $n/2 = 100/2 = 50$ $c = 4$
- $N_{j-1} = 49$
- $N_j = 74$ $Y_{j-1} = 38$

- $Me = Y_{j-1} + c \left[\frac{(n/2 - N_{j-1})}{(N_j - N_{j-1})} \right],$
- $= 38 + 4 \left[\frac{(50 - 49)}{(74 - 49)} \right] =$
- $= 38 + 4 \left[\frac{(1)}{(25)} \right] =$
- $= 38 + 4 \left[0,04 \right] =$
- $= 38 + 0,16 =$
- $= 38,16$

Podemos ahora decir que el 50% de los clientes tiene 38 años o menos y que el restante 50% tiene 38 años o más.

Veamos el gráfico de la variable edad:



3.1.2.4 Cuándo extraer la moda, la mediana y la media

Las medidas de posición sirven según sea el tipo de variable que queremos medir. Cuando tenemos variables cualitativas categóricas, la única medida de posición cuyo cálculo tiene sentido es la moda. Recordemos que la moda es la categoría de la variable con mayor frecuencia de aparición (Blanch y Joekes, 1997c).

Cuando por el contrario, nos encontramos con variables cuantitativas discretas podemos, además de extraer la moda, calcular la mediana para decir, por ejemplo, que la mitad de las observaciones tienen un valor igual o menor que el valor mediano y la otra mitad un valor igual o superior (Blanch y Joekes, 1997c).

Sólo para variables cuantitativas tiene sentido calcular las tres medidas de posición estudiadas (Blanch y Joekes, 1997c).

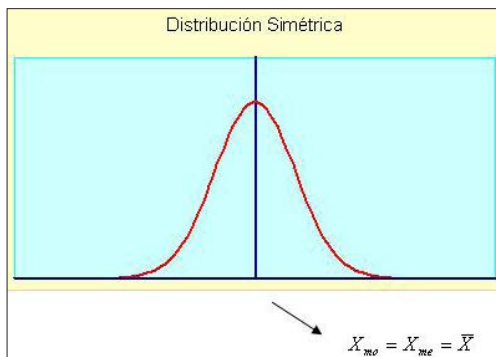
3.1.2.5 Utilidades de las medidas de tendencia central

Una manera útil de trabajar con datos estadísticos es realizando un análisis exploratorio a partir de la lectura de dos medidas importantes: la media y la mediana.

Si quisiéramos estudiar, por ejemplo, el promedio de ingresos en una población o muestra, y existieran pocas familias que ganan mucho y muchas familias que ganan poco, la media o promedio de ingresos de la población no va a resultar en una medida representativa puesto que su valor va a quedar afectado por los valores extremos de la variable (Blanch y Joeques, 1997c). La población, en este caso, no sería homogénea, habría mucha dispersión de los datos, es decir, mucha variabilidad. En estos casos es importante trabajar con la mediana, que divide al total de datos en dos mitades exactamente iguales sin quedar afectada por los valores extremos de la variable “Nivel de Ingresos”.

Es decir, la media aritmética no es una medida representativa de una serie de datos cuando existen valores atípicos (Blanch y Joeques, 1997c).

El modelo de distribución de frecuencias que se utiliza en Estadística presupone que los datos se distribuyen de manera homogénea debajo de una campana. Veamos el gráfico:



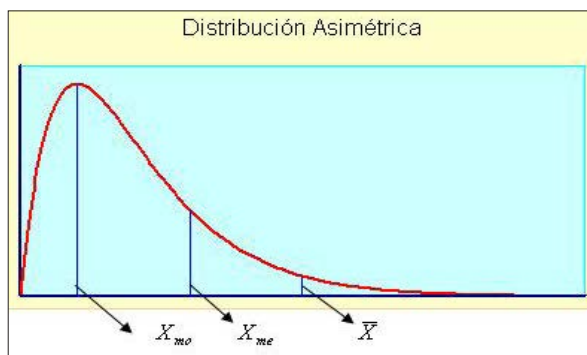
Según este modelo, los valores de una variable se distribuyen simétricamente alrededor de un valor central, que es el valor modal de la variable que coincide idealmente con la media y con la mediana.

Cuando esto sucede, decimos que la población es homogénea, que los datos son simétricos, es decir, que no existen valores extremos reales en una dirección particular, de forma tal que valores altos y bajos se compensan entre sí.

En la práctica concreta de la investigación estadística esto ocurre sólo en escasas oportunidades.

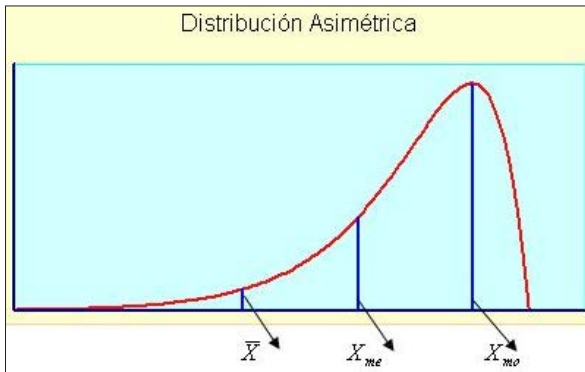
Generalmente, los conjuntos de observaciones no se encuentran simétricamente distribuidos, es decir, la distribución de los valores de la variable en estudio suele ser sesgada.

Se dice que el sesgo de la distribución es positivo cuando “la mayor proporción de las observaciones son menores o iguales que la media” (Blanch y Joeques, 1997e, 18), lo cual significa que la media adopta valores mayores a la mediana, debido a la presencia de escasas observaciones con valores extremadamente altos en el extremo derecho de la distribución (los cuales afectan su cálculo) (Blanch y Joeques, 1997e).



Si estuviéramos analizando un caso concreto y nos encontráramos con que la media es mayor que la mediana, esto querría decir que existen escasos valores extremos en la cola derecha de la distribución que compliquen el uso de la media como promedio general del total de la población (Blanch y Joekes, 1997e); (Triola, 2000); Berenson y Levine, 1996). Por ejemplo, podríamos sostener que existen pocos sujetos con ingresos muy altos y una concentración de ellos con ingresos bajos.

Mientras tanto, una distribución negativamente sesgada tiene más observaciones mayores o iguales a la media y presenta, en consecuencia, una mediana mayor que la media (Blanch y Joekes, 1997 e; Triola, 2000; Berenson y Levine, 1996). La campana de una distribución sesgada negativamente tiene una cola más larga a la izquierda de la distribución. Es sesgada hacia la izquierda por presentar escasos valores extremadamente bajos.



Si estuviéramos analizando un caso, diríamos que existen algunos valores extremos en la cola izquierda de la distribución. Probablemente tendríamos, siguiendo el ejemplo anterior, pocos casos con pocos ingresos que deformen la distribución y hagan que la media no sea una medida representativa de la serie de datos (por la presencia de los valores atípicos). En este caso, la media es menor que la mediana.

En suma, importante es conocer y diferenciar la media de la mediana puesto que esto nos permite tener una idea acerca de cómo se distribuyen los valores de la variable, cuál es la medida más útil y representativa de la serie de datos y el significado de tal distribución (Blanch y Joekes, 1997e; Triola, 2000; Berenson y Levine, 1996).

Ejemplo (adaptado de Berenson y Levine, 1996, 128).

En un estudio sobre el desempeño de alumnos de tres cursos de una escuela se encontró lo siguiente:

En un primer año se obtuvo un puntaje promedio de desempeño en la asignatura Literatura de 80 puntos. Mientras tanto, la mediana fue de 90 puntos. En este caso particular, la media es inferior a la mediana, o lo que es igual, la mayoría de los valores se concentra en el ala derecha de la distribución y existen escasos valores extremos en la cola izquierda que afecten el cálculo de la media. Se trata entonces de una distribución sesgada hacia la izquierda. En consecuencia, se puede establecer que el desempeño de los alumnos ha sido bueno (puesto que la verdadera concentración de valores está en la cola derecha).

En un segundo año, los alumnos, en la materia Literatura II, obtuvieron un promedio de 60 puntos. Mientras tanto, la mediana fue de 50 puntos. En este caso particular, la media es mayor que la mediana, lo cual significa que hay escasos valores altos en la cola derecha de la distribución. Lo anterior significaría que el rendimiento de los alumnos fue malo (puesto que hay mayor concentración de valores en la cola izquierda de la distribución).

En un tercer año, en la materia Literatura III, los alumnos obtuvieron un puntaje promedio de 70 y una mediana también de 70. En este grupo el desempeño ha sido homogéneo. Se trata de un conjunto de datos que se distribuyen de manera simétrica y que describen un desempeño *normal* del curso.

3.1.3 Medidas de dispersión

Reflejan la dispersión, oscilación de los datos, respecto a las medidas de tendencia central.

Complementan las de tendencia central para la descripción de una *distribución*.

3.1.3.1 Amplitud o rango

- Permite establecer la diferencia entre el valor más alto y más bajo de la distribución aunque ofrece poca información sobre la agrupación de los datos.
- Indica el *campo de variabilidad*.
- Suele acompañar a la moda.

3.1.3.2 Varianza

La varianza resulta en un estadístico que toma en consideración el modo en que fluctúan las observaciones alrededor de la media. La varianza de las observaciones es la suma de las diferencias cuadradas entre cada valor de la variable y la media aritmética, dividida por el tamaño de la muestra menos 1 (Berenson, Levine y Krehbiel, 2001, 115). La varianza es una medida razonablemente buena de la variabilidad de los datos (Canavos, 1999): si muchas de las diferencias son grandes, el valor de la varianza será grande; si, por el contrario, muchas de las diferencias son pequeñas, el valor de la varianza será pequeño. El valor de la varianza puede sufrir un cambio muy desproporcionado, aun más que la media, por la existencia de algunos valores atípicos extremos en el conjunto.

3.1.3.2.1 Fórmula para datos no agrupados

$$s = \frac{\sum (x_i - \text{media})^2}{n - 1},$$

3.1.3.2.1.1 Ejemplo

En una empresa de telefonía celular, tras la atención a 10 clientes en ventanilla, se decide estudiar la variabilidad en el tiempo de espera del cliente. Supongamos que los tiempos de espera fueron: 6,5 minutos; 6,6 minutos; 6,7 minutos; 6,8 minutos; 7,1 minutos; 7,3 minutos; 7,4 minutos; 7,7 minutos; 7,7 minutos; 7,7 minutos (Triola, 2000, 78).

Para hacer el cálculo se construye la siguiente tabla, tomada y adaptada de Triola (2000, 78).

Tiempo de espera de los clientes en la cola de la sucursal 1 de Telefonía Móvil

xi	xi - media	(xi - media) ²
6,5	6,5 - 7,15 = -0,65	0,4225
6,6	6,6 - 7,15 = -0,55	0,3025
6,7	6,7 - 7,15 = -0,45	0,2025
6,8	6,8 - 7,15 = -0,35	0,1225
7,1	7,1 - 7,15 = -0,05	0,0025
7,3	7,3 - 7,15 = 0,15	0,0225
7,4	7,4 - 7,15 = 0,25	0,0625
7,7	7,7 - 7,15 = 0,55	0,3025
7,7	7,7 - 7,15 = 0,55	0,3025
7,7	7,7 - 7,15 = 0,55	0,3025
Totales		2,045
71,5		
media=		s= $\sum (xi - media)^2 / n$
71,5/10=		
media= 7,15 minutos		s= 2,0450

3.1.3.2 Fórmula para datos agrupados

En el ejemplo anterior, la varianza se extrajo para datos no agrupados. Si se tratara de datos agrupados, la fórmula sería la siguiente:

$$s = \frac{\sum n_i (x_i - \text{media})^2}{n - 1},$$

dónde:

- n_i representa a la frecuencia absoluta de cada valor de la variable. Por ejemplo, si hubiéramos contabilizado en el estudio anterior los tiempos de espera de la totalidad de clientes a lo largo de una semana, tendríamos seguramente, muchos sujetos que tuvieron un tiempo de espera de 6,5 minutos; muchos sujetos con un tiempo de espera de 6,6 minutos, y así sucesivamente.

Además, si se decidiera generar intervalos de clase, el valor de x_i sería el punto medio de cada uno de los intervalos.

Dejemos la interpretación de la varianza en suspenso. Vayamos ahora a la definición de la desviación estándar.

3.1.3.3 Desviación estándar

La desviación estándar es la raíz cuadrada de la varianza, o lo que es igual, “raíz cuadrada de la suma del cuadrado de las diferencias con relación a la media aritmética, dividida entre el tamaño de la muestra menos 1” (Berenson, Levine y Krehbiel, 2001, 118).

3.1.3.3.1 Fórmula para datos no agrupados:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

3.1.3.3.2 Para datos agrupados:

$$S = \sqrt{\frac{\sum ni(x_i - \bar{x})^2}{n - 1}}$$

Importante

En la fórmula de la desviación estándar, el denominador es $n-1$ porque aquí se trabaja a nivel muestral y no poblacional. Si se quisiera extraer la desviación estándar poblacional, entonces el denominador sería N o total poblacional. Además, si se trabajara a nivel poblacional, en vez de utilizar la media muestral, tendríamos en la fórmula el signo μ , correspondiente con la media poblacional. Lo mismo se aplicaría a la fórmula de varianza poblacional.

En el caso del tiempo de espera de los clientes en la empresa de telefonía celular, la desviación estándar es igual a:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

$$S = \sqrt{2,0450}$$

$$S = 1,43 \text{ minutos}$$

La desviación estándar indica que la mayoría de los valores de la variable se agrupa dentro de 1,43 minutos alrededor de la media de 7,15 minutos, o lo que es lo mismo, entre:

$$\text{media} - 1s = 5,72 \text{ y } \text{media} + 1s = 8,58.$$

Para entender aún más la utilidad de la desviación estándar, tomemos por caso la variable “ingresos mensuales”. Si la desviación estándar de la variable “ingresos mensuales”, en una muestra o

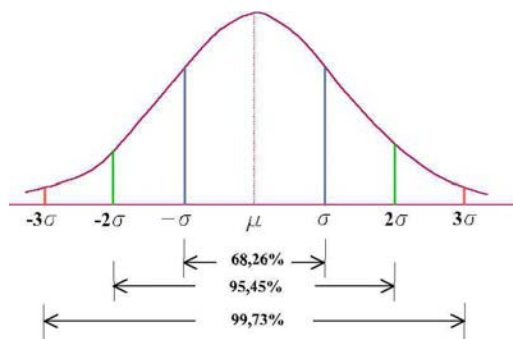
población determinada, es de \$100, esto quiere decir que la mayoría de los sujetos se agrupan en \$100 alrededor de la media. Si la media fuera, por ejemplo, de 1700 pesos, entonces podríamos decir que, la mayoría de los sujetos (un 68,26%) de la muestra o población mantienen ingresos de entre 1600 y 1800. Se sugiere leer la sección: 3.2.4.3.4 Utilidad de la desviación estándar para mayores detalles.

3.1.3.4 Uso de la varianza y la desviación estándar

La varianza y la desviación estándar miden la “dispersión promedio alrededor de la media” (Berenson, Levine y Krehbiel, 2001, 118), es decir, estiman la fluctuación de las observaciones que se encuentran por encima de ella y la de aquéllas que se encuentran por debajo de ella. Si bien la varianza es una medida útil, importante es recordar que su cálculo se expresa en unidades al cuadrado, edad al cuadrado, porcentajes al cuadrado, pesos al cuadrado, centímetros cuadrados, entre otros. Por lo tanto, para medir variación de los datos, será conveniente tomar en consideración a la desviación estándar, puesto que ésta se expresa en los valores originales de las variables: edad, porcentajes, centímetros, pesos.

3.1.3.5 Utilidad de la desviación estándar

Permite extraer intervalos dentro de los cuáles se localizan la mayoría de los valores observados (Triola, 2000; Berenson, Levine y Krehbiel, 2001; Berenson y Levine, 1996). El intervalo comprendido entre la media más/menos 1 desviación estándar incluye a una mayoría de los valores de los datos: generalmente, el 68,26% de los datos se encuentran entre $-1s$ y $+1s$. Luego, el 95,44% se ubica entre $-2s$ y $+2s$; y el 99,74% entre $-3s$ y $+3s$.



La desviación estándar permite comparar la variabilidad de los datos en muestras diferentes cada vez que las medias son iguales. Siguiendo el ejemplo anterior, podríamos querer comparar la desviación estándar del tiempo de espera de los 10 clientes de la compañía telefónica (de 1,43 minutos) con la desviación estándar del tiempo de espera de otros 10 clientes de otra empresa (de 1,82 minutos). Lo anterior significa que hay una variación mucho menor en cuanto al tiempo de espera de los clientes de la primera compañía en relación con el de aquellos correspondientes a la segunda.

3.1.3.6 Utilidad de las medidas de dispersión

Como puede observarse la varianza y la desviación estándar se utilizan para variables cuantitativas. Se trata de medidas de dispersión que permiten calcular el nivel de variación de los datos cuantitativos. Puede utilizarse para una sola población o muestra o para comparar los parámetros en varias (Triola, 2000; Berenson, Levine y Krehbiel, 2001; Berenson y Levine, 1996).

3.1.3.7 Coeficiente de variación

El coeficiente de variación es una indicación relativa de la variación de los datos (Triola, 2000; Berenson, Levine y Krehbiel, 2001; Berenson y Levine, 1996). Mide la dispersión porcentual de

los datos en relación con la media. El coeficiente de variación es igual a la división entre la desviación estándar y la media aritmética multiplicada por 100 (por ciento).

3.1.3.7 1 Fórmula

$$CV = (S/\text{media}) * 100\%$$

En el ejemplo del tiempo de espera de los clientes, el coeficiente de variación sería igual a:

$$CV = (1,43 / 7,15) 100\% = 20\%$$

Esto quiere decir que la variación de los tiempos de espera alrededor de la media es de un 20%.

3.2 Algunas estrategias de redacción de informes estadísticos

Para redactar informes estadísticos suelen utilizarse algunas estrategias universalmente aceptadas. A continuación se exponen algunas modalidades típicas de redacción; luego se presenta, a modo de ejemplo, un breve extracto de análisis. En el anexo se adjuntan tablas de frecuencias para que el principiante arme un informe propio.

- En relación a... podemos señalar que $\frac{1}{4}$ de los empleados (en vez de decir un 25%)...
- Más de la mitad (entre paréntesis escribir el porcentaje que debe ser superior al 50%).
- La mayoría de los encuestados... (utilizar la mayoría cuando el porcentaje es del 50% o más).
- La casi totalidad (cuando el porcentaje es cercano al 100%).
- A diferencia de... hallamos en este caso que el porcentaje de... es significativamente menor (cuando la diferencia entre

- un porcentaje y el otro es notable).
- La categoría con mayor frecuencia de aparición para la variable... es... con un porcentaje del...%.
 - Si bien la mayoría de los empleados, un... %, de un total de 138 son..., el porcentaje de ellos que..., aunque significativo, es levemente menor de lo esperado.
 - La media, de \$200, es más baja que la mediana, de \$400, hecho que indica que la librería xx ha trabajado de manera eficiente en el último mes. Existen escasos valores bajos en la cola izquierda de la distribución, hecho que afecta el cálculo de la media. La mediana es un valor más confiable para estimar el desempeño de la librería en un mes. El nivel de ganancia es alto.
 - La media, de 7 puntos en las notas obtenidas por los alumnos en el examen de Sociología, es mayor que la mediana, de 5 puntos, lo cual quiere decir que el grupo no logró un desempeño eficiente. La media queda afectada por los escasos valores altos en la cola derecha de la distribución, o lo que es lo mismo, existe concentración de valores en la cola izquierda, donde se encuentran la mayoría de los alumnos, los cuales sacaron notas bajas.

3.2.1 Un ejemplo de lectura de datos

La mayoría de los estudiantes de 1º año de la carrera de Comunicación de la UNC son de nacionalidad argentina: un 89% del total de 138. Mientras tanto, apenas un 2% es de origen europeo.

En términos de la cantidad de hombres y mujeres, hay predominio de varones. Ellos constituyen un 55% del total de estudiantes (más de la mitad) y ellas un 39% (porcentaje notablemente menor).

Cerca de la totalidad, (ó la casi totalidad), un 90% del total de 138 empleados, finalizó los estudios secundarios.

El grupo más numeroso, o cerca de la mitad de estudiantes no tiene estudios superiores (un 45%). Mientras tanto, un significativo

casi 30% (o el segundo grupo numeroso, un 30% de sujetos del total de 138 encuestados) finalizó estudios universitarios previamente al ingreso a la carrera o empezó y abandonó este tipo de estudios.

En relación con el estado civil, un 73%, o lo que es igual, cerca de $\frac{3}{4}$ de un total de... estudiantes, es soltero, mientras que el segundo grupo numeroso, aunque significativamente más pequeño, (apenas un 9%) es casado o tiene pareja estable.

3.3 Ejercicios

- 1) En el siguiente cuadro señale con una cruz en la columna correspondiente, el tipo de variable de que se trate.

Variable	Catagórica	Catagórica dicotómica	Cuantitativa discreta	Cuantitativa continua
Horas de trabajo semanales				
Edad				
Género				
Cantidad de promociones en el trabajo				
Calidad de atención del vendedor (muy buena, buena, regular, mala, muy mala)				
Nivel educativo				
Características de las relaciones de trabajo (muy buenas, buenas, regulares, malas, muy malas)				
¿Experiencia laboral previa? (sí o no)				
Probabilidad de ser ascendido en el trabajo (muy probable, probable, no está seguro, improbable, muy improbable)				

- 2) Supongamos que el Departamento de Marketing de una

- 3) En el presente ejercicio presentamos parte de un cuestionario tomado y adaptado de Hayes (1998).
 - i) Lea atentamente las preguntas que se exponen a continuación para saber qué deberá hacer en este ejercicio.
 - ii) Lea de manera completa el cuestionario.
 - iii) Responda las preguntas planteadas a continuación.

Preguntas

- a) Clasifique las variables en cualitativas y cuantitativas.
- b) Establezca el nivel de medición de cada una de las variables involucradas. En cada caso, justifique su respuesta.
- c) Determine en cada caso, el tipo de pregunta formulado.
- d) Determine la manera en que usted aplicaría el cuestionario.
- e) A partir de las preguntas 2, 4, 5, 6, 7 y 8 del cuestionario, deduzca qué tipo de target o público es el que interesa al Departamento de Marketing de la empresa (grupo etario, grupo social) y explique las razones.
- f) ¿Cuál es el objetivo del cuestionario? ¿Qué información busca el Departamento de Marketing de la empresa editorial?
- g) Qué otras preguntas formularía y dónde las ubicaría para averiguar además, si, a) los clientes consumen otros periódicos y/u otras publicaciones como revistas, y si los consumen, ¿cuáles y con qué frecuencia?; b) qué secciones del periódico les interesa más y cuáles de los diarios/periódicos/revistas de la competencia.
- h) ¿Cambiaría de lugar algunas preguntas del cuestionario? ¿Por qué?.

El cuestionario que se presenta a continuación fue tomado y adaptado de Hayes (1998,160-161).

1. Usted es: varón mujer no contesta
2. ¿Cuál es su edad? Menor de 18 30-34 45-49
 18 a 24 35-39 50-54
 25 a 29 40-44 55 o
más no contesta
3. ¿Cuál es su Estado civil? casado soltero divorciado
/ separado
 viudo unión de hecho
 no contesta

4. ¿Tiene usted empleo actualmente? sí no no
contesta

Si respondió sí, pasar a la pregunta 5. Si su respuesta es no, pasar a la pregunta 7.

5. ¿Es usted actualmente empleado full o part-time?
 empleado full time, esto es, 30 horas o más a la semana
 empleado part-time, esto es, menos de 30 horas a la
semana
 no contesta

6. Por favor, seleccione la categoría que mejor describa
su tipo de empleo/ ocupación (seleccione sólo una opción)

- educador
- trabajo en salud trabajo en construcción (como
mano de obra calificada: ingenieros, arquitectos)
- ama de casa estudiante
- desocupado ejecutivo/management
- empresario empleadopúblico/administrativo
- comerciante funcionario

- jubilado profesión liberal
 Otros. Especifique: _____ no contesta

7. ¿Cuál es el monto total de ingresos mensuales en su familia?

- entre 701 y 900
- entre 901 y 1100
- entre 1101 y 1300
- entre 1301 y 1500
- de 1500 a 2000
- de 2001 a 3000
- de 3001 a 4000
- de 4001 a 5000
- de 5001 a 6000
- más de 6000
- no contesta

8. ¿Cuál es el máximo nivel educativo que usted ha adquirido?

- secundario completo
- terciario incompleto
- terciario completo
- universitario incompleto
- universitario completo
- postgrado incompleto
- postgrado completo
- especialista incompleto
- especialista completo
- magíster incompleto
- magíster completo
- doctorado incompleto
- doctorado completo
- pos-doc incompleto
- pos-doc completo
- no contesta

9. ¿Con qué frecuencia lee usted el periódico XX?
 mensualmente cada 2 meses cada 1 mes y medio
 no contesta
10. Si es usted casado, o unido de hecho, su cónyuge ¿lee XX?
 sí no no contesta
11. Incluyéndose usted mismo, cuántos adultos en su familia leen usualmente el periódico XX?
 1 2 3 4 5 6 o más no contesta
12. Por favor, estime cuánto tiempo dedica usted a leer un número del periódico XX.
 15' aproximadamente 45' aproximadamente
 más de 1 hora
 30' aproximadamente 1 hora aproximadamente
 no contesta
13. Por favor, señale su grado de acuerdo con las siguientes afirmaciones respecto del periódico XX. Marque con una cruz el casillero que mejor refleje su opinión.

	Muy de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Muy en desacuerdo	No contesta
Ante todo, estoy satisfecho con la calidad de los artículos publicados por XX						
Los artículos en XX son de carácter informativo						
Disfruto leer XX						

4) Análisis de variables e informe estadístico

- 1.1) En función de las tablas de frecuencia que se muestran a continuación, completar el informe sobre la situación de los supuestos 138 alumnos ingresantes a 1° año de la carrera de Comunicación Social de la UNC.

Tablas de Frecuencia

	Frecuencia	Porcentaje	Porcentaje valido	Porcentaje acumulado
validos Argentina	123	89.9		
Chile	1	.7	95.3	
Otros Países	2	1.4	.8	95.3
Europa	3	2.2	1.6	96.1
Total	129	93.5	2.3	97.7
Perdidos Sistema	9	6.5	100.0	100.0
Total	138	100.0		

Cantidad de hombres y mujeres en 1° año de la carrera de Comunicación Social

	Frecuencia	Porcentaje	Porcentaje valido	Porcentaje acumulado
Validos Masculino	76	55.1	58.5	58.5
Femenino	54	39.1	41.5	100.0
Otros Países	130	94.2	100.0	
Total	8	5.8		
Perdidos Sistema	138	100.0		
Total				

Cantidad de sujetos que egresaron del secundario

	Frecuencia	Porcentaje	Porcentaje valido	Porcentaje acumulado
Validos si egresó	125	90.6	100.0	100.0
Perdidos Sistema	13	9.4		
Total	138	100.0		

Nivel de Estudios de los estudiantes de 1º año de la carrera de Comunicación Social

Estudios superiores		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	No tiene estudios superiores	62	44.9	50.8	50.8
	Continúa estudios terciarios	3	2.2	2.5	53.3
	Abandonó estudios terciarios	5	3.6	4.1	57.4
	Finalizó estudios terciarios	4	2.9	3.3	60.7
	Continúa estudios universitarios	25	18.1	20.5	81.1
	Abandonó estudios universitarios	15	10.9	12.3	93.4
	Finalizó estudios universitarios	8	5.8	6.6	100.0
	Total	122	88.4	100.0	
Perdidos	Sistema	16	11.6		
Total		138	100.0		

Estado civil de los alumnos de 1º año de la carrera de Comunicación Social

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Soltero/a	102	73.9	85.0	85.0
	Casado/a - Pareja Estable	13	9.4	10.8	95.8
	Viudo/viuda	2	1.4	1.7	97.5
	Divorciado/a	3	2.2	2.5	100.0
	Total	120	87.0	100.0	
Perdidos	Sistema	18	13.0		
Total		138	100.0		

¿Con quién vive actualmente?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Solo	23	16.7	19.2	19.2
	Con compañeros	15	10.9	12.5	31.7
	Con su pareja	14	10.1	11.7	43.3
	Con sus padres	46	33.3	38.3	81.7
	Con otros familiares	16	11.6	13.3	95.0
	Con su cónyuge	6	4.3	5.0	100.0
	Total	120	87.0	100.0	
Perdidos	Sistema	18	13.0		
Total		138	100.0		

¿Hijos?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	no	101	73.2	84.2	84.2
	uno	9	6.5	7.5	91.7
	dos	4	2.9	3.3	95.0
	más de dos	6	4.3	5.0	100.0
	Total	120	87.0	100.0	
Perdidos	Sistema	18	13.0		
Total		138	100.0		

¿Tiene obra social?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	no	44	31.9	36.7	36.7
	si, por su trabajo	14	10.1	11.7	48.3
	si, por un familiar	61	44.2	50.8	99.2
	No sabe/no contesta	1	.7	.8	100.0
	Total	120	87.0	100.0	
Perdidos	Sistema	18	13.0		
Total		138	100.0		

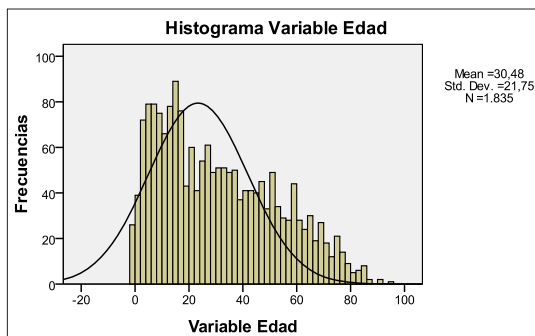
- 1.2) A continuación se presentan los datos obtenidos del análisis de la Encuesta Permanente de Hogares del año 2007 para el Aglomerado Concordia. Las variables que se trabajaron son: edad; sexo; nivel educativo; nivel ocupacional. En un primer término, el lector encontrará tablas y gráficos que describen a la variable “Edad”. En segundo término, encontrará la descripción de la variable “Sexo”. Luego, observará dos tablas de doble entrada. En la primera, se relacionan las variables “sexo” y “nivel educativo”. En la segunda, las variables “nivel educativo” y “nivel ocupacional”. Se pide que el estudiante realice:
- a) Un análisis detallado de la variable edad. Indicar nivel de medición de la variable; analizar grupos modales, media, mediana y desviación estándar. Determinar el tipo de distribución de que se trata y por qué. Describir el tipo de población.
 - b) Realizar un análisis de la variable “sexo”. Indicar nivel de medición de la variable.
 - c) Completar el informe sobre el cuadro que cruza las variables “sexo” y “nivel educativo”. Indicar el nivel de medición de las variables analizadas.
 - d) Completar el informe sobre el cuadro que cruza las variables “nivel educativo” y “nivel educacional”. Indicar el nivel de medición de las variables estudiadas.
 - e) Redactar una breve conclusión respecto de la información manejada en los análisis anteriores.

Tablas de frecuencia/contingencia y gráficos

Variable "Edad"

Concordia	Total de sujetos	Casos Válidos	1835
	(N)	Casos Perdidos	0
	MEDIA		30,48
	MEDIANA		27,00
	MODA		15
	DESVIACIÓN ESTÁNDAR		21,750
	VARIANZA		473,054

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Entre -1 y 20 años	764	41,6	41,6	41,6
Entre 21 y 40 años	474	25,8	25,8	67,5
Entre 41 y 60 años	384	20,9	20,9	88,4
Entre 61 y 80 años	193	10,5	10,5	98,9
Entre 81 y 99 años	20	1,1	1,1	100,0
Total	1835	100,0	100,0	



Sexo	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Varón	892	48,6	48,6	48,6
Mujer	943	51,4	51,4	100,0
Total	1835	100,0	100,0	

Nivel Educativo según Sexo

Aglomerado Concordia	Nivel Educativo							Total
	Primaria Incompleta	Primaria Completa	Secundaria Incompleta	Secundaria Completa	Superior Universitaria Inc.	Sup. Univ. Comp.	Sin Instrucción	
	233	151	197	106	41	44	120	892
% según sexo	26,1%	16,9%	22,1%	11,9%	4,6%	4,9%	13,5%	100,0%
Varón % según nivel educativo	51,2%	44,2%	52,4%	52,0%	42,7%	34,1%	51,5%	48,6%
% según total	12,7%	8,2%	10,7%	5,8%	2,2%	2,4%	6,5%	48,6%
	222	191	179	98	55	85	113	943
% según sexo	23,5%	20,3%	19,0%	10,4%	5,8%	9,0%	12,0%	100,0%
Mujer % según nivel educativo	48,8%	55,8%	47,6%	48,0%	57,3%	65,9%	48,5%	51,4%
% según total	12,1%	10,4%	9,8%	5,3%	3,0%	4,6%	6,2%	51,4%
	455	342	376	204	96	129	233	1835
% según sexo	24,8%	18,6%	20,5%	11,1%	5,2%	7,0%	12,7%	100,0%
% según nivel educativo	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
% según total	24,8%	18,6%	20,5%	11,1%	5,2%	7,0%	12,7%	100,0%

En relación con el cuadro anterior se puede señalar que:

- El ... de un total de 892 varones tiene primaria incompleta o primaria completa. Significativamente, de un total de 943 mujeres un porcentaje igual (con una diferencia mínima de un punto, un....) mantiene este mismo nivel de estudios.
- Un ... de varones, mantiene estudios secundarios incompletos o estudios secundarios completos. En el grupo de mujeres este porcentaje disminuye a un
- Mientras un ... de varones presenta estudios superiores incompletos o completos; el porcentaje de mujeres que ostenta este nivel de estudios es de un Las mujeres presentan ventajas en relación con los varones a este nivel de estudios.

Llama la atención que cerca de la mayoría de los varones y de las mujeres del aglomerado Concordia apenas tengan estudios primarios (sean éstos completos o incompletos). A lo anterior hay que agregar que un de los varones y un 12% de las mujeres censadas/os no tiene ningún nivel de instrucción.

Estado ocupacional según nivel educativo

Aglomerado Concordia		Estado ocupacional				Total	
		Ocupado	Desocupado	Inactivo	Menor de 10 años		
nivel ed	Prim. Incomp.	Cantidad	69	10	213	163	455
		% según nivel educativo	15,2%	2,2%	46,8%	35,8%	100,0%
		% según estado ocupac.	10,3%	16,9%	29,1%	44,1%	24,8%
		% of Total	3,8%	5%	11,6%	8,9%	24,8%
	Prim. Comp.	Cantidad	182	13	147	0	342
		% según nivel educativo	53,2%	3,8%	43,0%	,0%	100,0%
		% según estado ocupac.	27,0%	22,0%	20,1%	,0%	18,6%
		% of Total	9,9%	7%	8,0%	,0%	18,6%
	Sec. Incomp.	Cantidad	126	15	235	0	376
		% según nivel educativo	33,5%	4,0%	62,5%	,0%	100,0%
		% según estado ocupac.	18,7%	25,4%	32,1%	,0%	20,5%
		% of Total	6,9%	8%	12,8%	,0%	20,5%
	Sec. Comp.	Cantidad	127	12	65	0	204
		% según nivel educativo	62,3%	5,9%	31,9%	,0%	100,0%
		% según estado ocupac.	18,9%	20,3%	8,9%	,0%	11,1%
		% of Total	6,9%	7%	3,5%	,0%	11,1%
Sup. Un. Inc.	Cantidad	45	8	43	0	96	
	% según nivel educativo	46,9%	8,3%	44,8%	,0%	100,0%	
	% según estado ocupac.	6,7%	13,6%	5,9%	,0%	5,2%	
	% of Total	2,5%	4%	2,3%	,0%	5,2%	
Sup. Univ. Comp.	Cantidad	110	1	18	0	129	
	% según nivel educativo	85,3%	,8%	14,0%	,0%	100,0%	
	% según estado ocupac.	16,3%	1,7%	2,5%	,0%	7,0%	
	% of Total	6,0%	,1%	1,0%	,0%	7,0%	
Sin Inst.	Cantidad	14	0	12	207	233	
	% según nivel educativo	6,0%	,0%	5,2%	88,8%	100,0%	
	% según estado ocupac.	2,1%	,0%	1,6%	55,9%	12,7%	
	% of Total	,8%	,0%	,7%	11,3%	12,7%	
Total	Cantidad	673	59	733	370	1835	
	% según nivel educativo	36,7%	3,2%	39,9%	20,2%	100,0%	
	% según estado ocupac.	100,0%	100,0%	100,0%	100,0%	100,0%	
	% of Total	36,7%	3,2%	39,9%	20,2%	100,0%	

Quando se cruzan las variables “estado ocupacional” según “nivel de educación”, se encuentran los siguientes valores:

- De un total de 455 sujetos censados que no han terminado el nivel primario de instrucción, cerca de la mitad (un.....) se encuentra inactivo. Sólo un de este grupo se encuentra ocupado.
- De un total de 342 sujetos encuestados que mantienen estudios primarios completos, la mayoría (un) está ocupado. Mientras tanto, la casi otra mitad (un) se encuentra inactivo.
- De 376 sujetos con estudios secundarios incompletos, más de la mitad (un) se encuentra inactivo. Mientras tanto, un se encuentra ocupado.
- Del total de 204 individuos que lograron finalizar los estudios secundarios, un se encuentra activo al momento del censo. Ahora bien, un porcentaje menor aunque no menos significativo (un) se encuentra inactivo.
- En cuanto al grupo minoritario y privilegiado de los sujetos que mantienen estudios universitarios incompletos, puede mencionarse que, del total de 96, se encuentran activos. Ahora bien, se encuentran inactivos.
- Finalmente, sólo 129 sujetos del total de 1835 censados del aglomerado de Concordia completaron estudios universitarios. De este grupo minoritario y privilegiado de 129 sujetos, un se encuentra activo. Sólo un inactivo.

Bibliografía

- Apezechea, H.M. (1979) “La codificación” en *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. México: FCE. Pp. 122-153
- Baranger, D. (1992) *Construcción y análisis de datos*. Posadas: Editorial Universitaria.
- Berenson, M. L. y D. M. Levine (1996) *Estadística Básica en Administración, conceptos y aplicaciones*. Sexta edición. México: Prentice Hall Hispanoamericana, S.A.
- Berenson, M. L., D. M. Levine y T.C. Krehbiel (2001) *Estadística para Administración*. Segunda Edición. México: Prentice Hall.
- Blalock, H. M. (1996) *Introducción a la Investigación Social*. Buenos Aires: Amorrortu editores.
- Blanco, C. (2007) “El perfil social y económico de los estudiantes-ingresantes a cuatro carreras de la Facultad de Filosofía y Humanidades de la UNC del año 2004”, Artículo publicado en Actas de las V Jornadas de Investigación en Educación. ISBN 978-950-33-0604-8. Córdoba, 4 y 5 de julio de 2007.
- Blanco, C. (2007) Informe sobre las encuestas realizadas en el marco de la VI Bienal Iberoamericana de Comunicación. Comunicación y poder. Movimientos sociales y medios en la consolidación de las democracias, Córdoba, Argentina, 26 al 29 de septiembre de 2007.
- Blanch N. y S. Joekes (1997b) “Organización y resumen de datos”. Módulo II del Curso de Postgrado Estadística Aplicada a la Investigación, Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.
- Blanch N. y S. Joekes (1997c) “Medidas de Posición”. Módulo III del Curso de Postgrado en Estadística Aplicada a la Investigación, Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.

- Blanch N. y S. Joekes (1997d) “Medidas de Dispersión”. Módulo IV del Curso de Postgrado en Estadística Aplicada a la Investigación, Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.
- Blanch N. y S. Joekes (1997e) “Nuevas Técnicas de Análisis”. Módulo V del Curso de Postgrado en Estadística Aplicada a la Investigación, Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.
- Blanch N. y S. Joekes (1997g) “Diseño de experimentos y selección de muestras aleatorias de aplicaciones finitas”. Módulo XI del Curso de Postgrado Estadística Aplicada a la Investigación, Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.
- Blanch N. y S. Joekes (1997f) “Inferencia Estadística. Estimación de parámetros y prueba de hipótesis”. Módulo X del Curso de Postgrado Estadística Aplicada a la Investigación, Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.
- Blanch, N. y S. Joekes (1997a) “El rol de la estadística en el método científico”. Módulo I del Curso de Posgrado “Estadística Aplicada a la Investigación”, del Departamento de Educación a Distancia, Facultad de Ciencias Económicas, UNC.
- Boudon R. y P. Lazarsfeld (1973). *Metodología de las Ciencias Sociales*. Barcelona: Editorial Laia. Tomo I.
- Bourdieu, P. (1973) *L'opinion publique n'existe pas*, en “Les temps modernes” n. 318 (janvier):1292-1309.
- Bourdieu, P. (1987) “Fieldwork in Philosophy”, en Bourdieu, P., *Choses Dites*. París: Ed. De Minuit. Pp. 13-46. (“Fieldwork in Philosophy”, en *Cosas Dichas*. Buenos Aires: Gedisa. 1988. Pp. 17-43)
- Bourdieu, P. (2006) *La distinción. Criterios y bases sociales del gusto*. 3º Edición. Madrid: Editorial Taurus.
- Bourdieu, P. (2007) *Campo del poder y reproducción social: elementos para un análisis de la dinámica de las clases*. 1º Edición. Córdoba: Ferreira Editor.
- Bourdieu, P., J.C. Chamboredon y J.C. Passeron (2004) *El oficio del sociólogo. Presupuestos epistemológicos*. 1º Edición. 1º Reimpresión. Buenos Aires: Siglo XXI Editores Argentina.
- Camacho Rosales, J. (2000) *Estadística con SPSS para Windows. Versión 9*. Madrid: Ra-Ma Editorial.

- Campbell, A. A. y Katona G. (1979) “La encuesta por muestreo: una técnica para la investigación en Ciencias Sociales” en *Los Métodos de Investigación en las Ciencias Sociales*. Buenos Aires: Paidós. Pp 31-66.
- Canavos, G.C. (1999) *Probabilidad y Estadística. Aplicaciones y Métodos*. México: Mc. Graw Hill.
- Cannell, C. F. y R. L. Kahn (1979) “La reunión de datos mediante entrevistas” en *Los Métodos de Investigación en Ciencias Sociales*. Buenos Aires: Paidós. Pp.310-352.
- Cartwright, D. P. (1979) “Análisis del material cualitativo” en *Los métodos de investigación en las Ciencias Sociales*. Buenos Aires: Paidós. Pp. 389-430.
- Corbetta, P (2007) *Metodología y Técnicas de Investigación Social*. Edición Revisada. Madrid: Mc. Graw Hill.
- Costa, R. (1976) *Para una definición social de los actores sociales. La teoría del campo*. Córdoba: Mimeo.
- Costa, R. (1986) *Recuperación del actor social y explicación sociológica de las prácticas sociales*. Córdoba: Editorial Mimeo.
- Costa, R. (2004) “Teoría y compromiso en Pierre Bourdieu. Mecanismos sociales del cambio”. Documento de trabajo. Primera edición. Córdoba: Editorial de la Universidad Católica de Córdoba.
- De Luque, S. (1997) “La problemática valorativo-metodológica de las Ciencias Sociales” en Díaz E. (Ed.). *Metodología de las Ciencias Sociales*. Buenos Aires: Editorial Biblos.
- Dorado, C. (2009). “Capítulo V: Encuesta” en *Herramientas de Metodología para investigar en comunicación*. 1º Edición. Córdoba: Copy-rápido. Tomo 2.
- Dorado, C. R. (2002) “Encuesta” en apunte de cátedra del Taller de Metodología de la Investigación Aplicada, parte 2. Pp. 57-71.
- Emanuelli, P. (2002) “Elaboración de hipótesis y determinación de las unidades de análisis” y “Unidades de análisis u observación” en apunte de cátedra del Taller de Metodología de la Investigación Aplicada, Unidad III. Pp. 66-106.
- Emanuelli, P. (2009) “Acerca de las hipótesis y los elementos que las constituyen” en *Herramientas de Metodología para investigar en comunicación*. 1º Edición. Córdoba: Copy-rápido. Tomo 1.

- Gallart, M.A. (1993) "La integración de métodos y la metodología cualitativa", en Forni, Gallart y Vasilachis: *Métodos cualitativos II: La práctica de la investigación*. Buenos Aires: Centro Editor de América Latina. Pp. 107-151.
- Geertz, C. (2003) *La interpretación de las culturas*. Madrid: Gedisa Editorial.
- Ghiglione, R. (1989) "Encuestar". En *Técnicas de investigación en Ciencias Sociales. Datos, observación, entrevista, cuestionario*. Blanchet, Ghiglione, Massonnat y Trognon Eds. Madrid: Nancea Editorial.
- Glasser, B.G. y A.L. Strauss (1967) *The Discovery of grounded theory*. New York: Aldine Publishing Company.
- Grasso, L. (1999) *Introducción a la Estadística en Ciencias Sociales y del Comportamiento*. Córdoba: Facultad de Psicología, UNC.
- Greenwood, E. (1973) *Metodología de la Investigación Social*. Buenos Aires: Editorial Paidós.
- Guell A. M. (1973). "Hipótesis y variables" en *Metodología de las Ciencias Sociales*. Barcelona: Edit. LAIA. Tomo I. pp. 47-62.
- Gutiérrez, A.B. (2005) *Las prácticas sociales: una introducción a Pierre Bourdieu*. Córdoba: Ferreyra Editor.
- Hayes, B. E. (1998) *Measuring Customer Satisfaction. Survey Design, use, and Statistical Analysis Methods*. Milwaukee Wisconsin: ASQ Quality Press.
- Korn, F. (1969) "El significado del término "variable" en sociología". En *Conceptos y variables en la investigación social*. Buenos Aires: Ed. Nueva Visión. Pp. 9-19.
- Lasswell, H. (1955) "Estructura y función de la comunicación en la sociedad", en Moragas M., *Sociología de la comunicación de masas*. Barcelona: Editorial Gustavo Gili. (1979) 1982.
- Lazarsfeld, P. (1969). "Nacimiento y desarrollo de las variables" en *Conceptos y variables en la investigación social*. Buenos Aires: Ed. Nueva Visión. Pp. 23-47.
- Lazarsfeld, P. (1973). "De los conceptos a los índices empíricos" en *Metodología de las Ciencias Sociales*. Boudon, R. y Lazarsfeld, P. Comp. Barcelona: Edit. LAIA. Tomo I. pp. 35-46.

- Le Compte, M.D. y J. Preissle (1996) "Toward and ethnology of student life in schools and classrooms: Synthesizing the qualitative research tradition. En M.D. Le Compte, W.L. Millroy y J. Preissle (Eds.) *The Handbook of qualitative research in education*. (págs.. 815-859). New York: Academic Press.
- Lininger, C.A. y D. Warwick (1978) *La Encuesta por Muestreo: Teoría y Práctica*. México: Compañía Editorial Continental S.A.
- Maintz, R., K., Holm y P. Hübner. (1992). *Introducción a los métodos de la sociología empírica*. Buenos Aires: Editorial Alianza.
- Marradi, A., N. Archenti, J.I. Piovani (2010) *Metodología de las Ciencias Sociales*. 1º Edición. Buenos Aires: Cengage Learning.
- Martínez Miguelez, M. (1999) "Criterios para la Superación del Debate Metodológico Cuantitativo/Cuantitativo" en Revista Interamericana de Psicología/Interamerican Journal of Psychology. Vol.33. Nº 1. Pp. 79-107.
- Mattelart, A. (1996) Capítulo 2: "La era de las multitudes" y capítulo 3 "La gestión invisible de la gran sociedad", en *La comunicación- mundo*. México: Siglo XXI Editores.
- Mattelart, A. y M. Mattelart (1997) *Historias de las Teorías de la Comunicación*. 1º Edición. Barcelona: Ed. Paidós
- Maxwell McCombs, Donald Shaw, 'The Agenda-Setting Function of Mass-Media', in *The Public Opinion Quarterly*, Vol. 36, No. 2. (Summer, 1972), pp. 176-187.
- Maxwell, J.A. (1996) *Qualitative Research Design. An Interactive Approach*. Londres: Sage. Traducción: Lic. María Luisa Graffigna. Revisión y corrección de la traducción: Dra. Irene Vasilachis de Gialdino.
- Meichsner, S. (2007) El campo político en la perspectiva teórica de Bourdieu. *Revista de Ciencias Sociales de la Universidad Iberoamericana (versión electrónica)*, Nº3, (Año 2, Otoño), pp.1-22. Departamento de Ciencias Sociales y Políticas, México.
- Morley D. y R. Silverstone (1993) "Públicos mediáticos. Comunicación y contexto: La perspectiva etnográfica en los sondeos de opinión" en *Metodologías cualitativas de investigación en comunicación de masas*. Jensen y Jankowski Eds. Primera Edición. Barcelona: Editorial Bosch.
- Padua, J. e I. Ahman (1979) "El cuestionario" en *Técnicas de investigación aplicada a las Ciencias Sociales*. México: FCE. Pp. 86-121.

- Peak, H. (1979) "Los problemas de la observación objetiva" en *Los métodos de investigación en las Ciencias Sociales*. Buenos Aires: Paidós. Pp. 235-285.
- Quivy R. y Van Compenhoudt (1998) *Manual de Investigación en Ciencias Sociales*. México: Editorial Limusa, S.A.
- Ramos. J. (1996) Apunte de la cátedra Teoría del Conocimiento y Lógica, de la Escuela de Ciencias de la Información, de la UNC.
- Rodríguez Gómez, G., Flores J.G. y Jiménez E.G. (1996) *Metodología de la investigación cualitativa*. Málaga: Ediciones Aljibe, S.L.
- Sampieri R.H., Collado F.C. y Lucio P.B. (2000) *Metodología de la Investigación*. Buenos Aires: Mc. Graw Hill. Segunda Edición.
- Sautu, R., P. Boniolo, P. Dalle y R. Elbert (2005) *Manual de Metodología. Construcción del marco teórico, formulación de objetivos y elección de la metodología*. Buenos Aires: CLACSO Editorial.
- Scheaffer, R. L., W. Mendenhall y Ott (1987) *Elementos de Muestreo*. México: Grupo Editorial Iberoamérica.
- Sierra Bravo, R. (1988) *Técnicas de Investigación Social*. Madrid: Paraninfo.
- Strauss, A. L. (1989) *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.
- Strauss, A. y J. Corbin (1990) *Basics of Qualitative Research. Grounded Theory, procedures and techniques*, Londres: Sage.
- Strauss, A. y J. Corbin (1994) "Grounded Theory Methodology. An Overview" en Denzin, N. K. y Lincoln, Y.S. (eds.): *Handbook of Qualitative Research*. California: Sage Publications.
- Taylor S.J. y R. Bogdan (1982) *Introducción a los métodos cualitativos de investigación*. Buenos Aires: Paidós.
- Triola, M.F. (2000) *Estadística Elemental*. Séptima Edición. México: Addison Wesley Longman.
- Vasilachis (1992) *Métodos cualitativos I*. Buenos Aires: CEDAL.
- Weiers, R. M. (1991) *Investigación de Mercados*. México, Englewood Cliffs, Nueva Delhi, Wellington, Londres, Río de Janeiro, Sydney, Singapur, Tokio, Toronto: Prentice Hall- Hispanoamericana S.A.

Wimmer y Dominick (1996) *La investigación científica de los medios de comunicación*. Primera Edición Española. Barcelona: Ed. Bosch.

Wolf, M. (1996) *La investigación de la comunicación de masas*. México: Editorial-Paidós.

Encuesta y Estadística

Métodos de Investigación Cuantitativa en Ciencias Sociales y Comunicación



Este libro fue pensado para todos aquellos profesionales, estudiantes, docentes e investigadores de las Ciencias Sociales y la Comunicación Social que requieran de las herramientas necesarias para llevar a cabo un proceso completo de aplicación del método de encuesta. Se desarrollan a lo largo de la obra, las implicancias tanto del método de encuesta como de la estadística. Se trata de una propuesta pedagógica basada en los aportes brindados por diferentes metodólogos, sociólogos, estadísticos y comunicólogos.

En consonancia con lo planteado por otros investigadores, se ubica a la encuesta en el marco más amplio de la metodología de la investigación social; los sondeos de opinión, la teoría estadística y los métodos de selección de muestras. Se concibe a la encuesta como una etapa posterior a la definición teórica del objeto de estudio. En este sentido, el libro ubica a la encuesta en el marco de un proceso de investigación, sea éste de carácter académico ó institucional, y pone énfasis en la definición conceptual del objeto de estudio como condición necesaria para la posterior operacionalización de variables. Se entiende que sin el procedimiento de operacionalización de variables no es posible establecer los datos pertinentes que permitirán luego construir preguntas acordes para un cuestionario.

