

LOS MÉTODOS, PROCEDIMIENTOS Y TÉCNICAS CUANTITATIVOS DE CONSTRUCCIÓN Y VALIDACIÓN DEL DISCURSO CIENTÍFICO.

Universidad de Buenos Aires Facultad de Ciencias Sociales Teórico N° 7
Carrera de Ciencias de la Comunicación

María Mercedes Di Virgilio

Materia: Metodología y técnicas de la investigación en Ciencias Sociales.
Cátedra: Prof. Salvia

Lecturas recomendadas del Teórico V:

LAZARSFELD, PAUL: "De los conceptos a los índices empíricos", en Raymond Boudon y Paul Lazarsfeld; Metodología de las ciencias sociales, Ed. Laia, Barcelona, 1973, Vol. I.

CORTES F. Y RUBALCAVA M. R., "Introducción." En Métodos Estadísticos Aplicados a la Investigación en ciencias Sociales. El Colegio de México, 1997.

ABRITTA, GUILLERMO: "Noción y Estructura del Dato". Publicación interna de la Cátedra. U.B.A. Facultad de Ciencias Sociales. Carrera de Comunicación. 1999.

RITCHEY, F., Estadística para ciencias sociales. El potencial de la imaginación estadística. Cap. 4, pag. 100 a 111. McGrau-Hill. Buenos Aires. 2001.

ARIAS ASTRAL, A. y FERNÁNDEZ RAMÍREZ, B., "La encuesta como técnica de investigación social". En Rojas Tejada, A., Fernández Prado, J. y Perez Meléndez, C. (Eds); Investigar mediante encuestas. Fundamentos teóricos y aspectos prácticos. Síntesis Psicología. Madrid. 1998

En las próximas páginas vamos avanzar en el conocimiento de algunos de los procedimientos que el investigador utiliza cuando desea llevar adelante una investigación empírica. Tal como ustedes han visto en los teóricos anteriores, en ciencias sociales se utilizan diferentes estrategias para llevar adelante procesos de investigación: las denominadas estrategias cuantitativas y cualitativas. Nos vamos a centrar hoy en *los procedimientos que involucra el desarrollo de una investigación cuantitativa*. Con el objetivo de poder "ver" concretamente cómo se hace una investigación con estas características vamos a recurrir al trabajo que desarrollaron Landi, Vacchieri y Quevedo sobre *Públicos y consumos culturales de Buenos Aires*. Vamos a recurrir a ese texto para ejemplificar la lógica de la investigación cuantitativa y para aprender algunas herramientas, como por ejemplo el uso de cuadros de contingencia para el análisis de datos producidos a través de encuestas.

Sin embargo, antes avanzar sobre la lógica con base en la cual se desarrolla la investigación cuantitativa, creo que nos puede ayudar hacer una breve revisión del proceso de investigación y de algunos elementos involucrados en su desarrollo que es necesario tener muy en claro antes de avanzar.

Siguiendo a Samaja (2001), podemos pensar el proceso de proceso de investigación como un conjunto de acciones/ actividades/procedimientos que desarrolla el investigador con vistas a responder preguntas, examinar ideas y/o poner a prueba teorías, es decir, con vistas a producir un conocimiento nuevo.

Tal como mencionáramos en el teórico III y IV, un elemento central de todo proceso de investigación es la *pregunta de investigación* ¿Qué es una pregunta de investigación? Es una pregunta que se responde con base en desarrollo de una

investigación empírica. Es decir, un tipo de investigación que incorpora en su desarrollo una combinación de *componentes teóricos* y *componentes empíricos* (Samaja, 2001).

Para responder estas preguntas, necesitamos recurrir a la evidencia empírica; no podemos basarnos exclusivamente en juicios morales, en preferencias o en puntos de vistas individuales sino que necesitamos construir una base empírica que nos permita avanzar en dar respuesta a nuestra pregunta.

En la investigación de Landi, Vacchieri y Quevedo, ellos parten de una pregunta general que alude a *¿cuál es el perfil de los consumos culturales de la Ciudad de Buenos Aires y del GBA?*¹ Los autores plantean que existe un supuesto aceptado acerca de que la Ciudad de Buenos Aires posee una rica tradición cultural. Sin embargo, sostienen que estas tradiciones se encuentran en un acentuado proceso de redefinición debido, por un lado, ante la grave crisis por la que atraviesan vastos sectores de la industria cultural y, por el otro, debido a la expansión de nuevas tecnologías comunicativas. Es en el marco de estas transformaciones que ellos se proponen *describir y analizar el perfil de los consumos culturales*. Es decir, *describir y analizar los hábitos/ comportamientos y gustos de porteños y bonaerenses de manera tal de poder construir un mapa de los públicos de las ofertas comunicativa y cultural*.

Como ustedes pueden notar, las preguntas de investigación se expresan en términos de una *relación*. Esta relación puede ser entre dos o más atributos de individuos o grupos, como por ejemplo, clase social y comportamiento electoral, género y movilidad social, exposición a la publicidad y efectos sobre los televidentes, posicionamiento político ideológico de los medios y formas de construcción de la noticia. La búsqueda y la postulación de relaciones entre atributos o características de individuos y grupos es un elemento central para avanzar en el conocimiento de la realidad social.

Como ya les mencionara en el Teórico IV, en ese proceso de conocimiento de la realidad social, *la teoría* nos brinda una gran ayuda, en la medida en que provee explicaciones sobre la relación que existe entre uno o más atributos de individuos y grupos. Es precisamente la teoría la que nos permite establecer los links entre lo que nosotros observamos (la base empírica) y nuestra perspectiva conceptual acerca de por qué ciertos fenómenos están vinculados con otros de una manera particular.

Por ejemplo, en la investigación de Landi, Vacchieri y Quevedo (1990), los autores desean comprender las diferencias entre los públicos y las pautas de consumo que caracterizan a cada uno; para avanzar en esa comprensión ellos postulan que determinados atributos de los individuos - nivel socio económico, nivel educativo, lugar de residencia - son los que contribuyen a delimitar públicos y consumos culturales diferenciales. De este modo, los públicos y los consumos culturales se encuentran diferenciados por razones económicas, de competencia educativa de los entrevistados y, también, por una diversidad regida por distintos gustos, tradiciones, gramática de desciframiento e identidades que responden a la heterogeneidad propia de la cultura popular (Landi et al, 1990:4 ss).

Esta proposición acerca de la manera en que los atributos observables en los individuos (nivel socioeconómico, nivel educativo, lugar de residencia) están relacionados con las pautas de consumo culturales en la vida real es una *hipótesis*. Recuerden que una hipótesis es una proposición que expresa una relación entre atributos o características de individuos o grupos. Es una respuesta tentativa a una

¹ En el desarrollo del trabajo plantean otra pregunta se refiere a *¿Cómo las pautas de consumo cultural definen/ inciden/ modelan las relaciones entre la escuela y los medios de comunicación -- en particular la TV --?*

pregunta de investigación: en este caso, a la pregunta acerca de cuáles son los atributos y/o características de los individuos que permiten construir un mapa de los públicos de las ofertas comunicativa y cultural.

En general, *las hipótesis se inscriben en un corpus teórico más amplio*. En el caso de Landi et al (1990), ese corpus teórico aparece algo velado. Sin embargo, algunos conceptos utilizados en la investigación sugieren algunas inscripciones; por ejemplo, el tratamiento de los objetos culturales (incluso de las obras de arte) como *bienes producidos en un circuito industrial* evoca las ideas de la Escuela de Frankfurt - ustedes han leído y analizado el texto de Adorno que es un exponente de esta corriente de pensamiento. Asimismo, la idea de que “ciertas manifestaciones culturales perviven en el tiempo en un permanente proceso de recreación histórica” (Landi et al, 1990:4), sugiere cierta afinidad con la *teoría del círculo hermenéutico* de Gadamer, según la cual las obras artísticas y culturales se van recreando y enriqueciendo a lo largo de la historia, en una relación iterativa entre pasado y presente, que incorpora nuevos elementos a la interpretación y al significado del objeto.

Sin embargo, no todas las hipótesis se derivan directamente de teorías,. Nosotros podemos generar hipótesis de muchas maneras: de la teoría, directamente de la observación, o a partir de la intuición. Probablemente la fuente más importante para la generación de hipótesis es la bibliografía sobre un campo específico de conocimiento. Una revisión crítica de la bibliografía existente, en general, no permite familiarizarnos con el estado del conocimiento en un campo de problemas y con las hipótesis que están siendo estudiadas.

Volvamos ahora a nuestra hipótesis:

Los públicos y las pautas culturales se encuentran diferenciados por razones económicas, de competencia educativa de los entrevistados y, también, por una diversidad regida por distintos gustos, tradiciones, gramática de desciframiento e identidades.

Nótese que esta hipótesis es una proposición que postula una relación entre *atributos y características de los individuos que son susceptibles de variar*: nivel socio económico, pautas culturales, nivel de instrucción. Estas características/ atributos de individuos o grupos se denominan *variables*. Por ejemplo, una persona puede ser clasificada según su nivel socio económico en nivel alto, nivel medio o nivel bajo. De manera similar, las personas tienen diferentes niveles de instrucción, es por ello que el nivel de instrucción es una variable, *puede asumir distintos valores* en cada individuo (vea la Tabla 1 con ejemplos de variables y sus posibles valores).

Tabla 1
Variables y sistemas de categorías

Variable	Categoría
Nivele socio económico	Nivel alto Nivel medio Nivel bajo
Tipo de revista que lee	Humor Deportes Historietas Femenina De actualidad Sensacionalista
Género	Masculino Femenino
Ingreso mensual	\$ 600 \$ 1.200 \$ 10.000 \$ 12.000
Tipos de medios de comunicación	Radio Diarios Televisión Revistas

Cada variable puede asumir diferentes valores. Al conjunto de valores de la variable se lo denomina *sistema de categorías*; las categorías deben ser *exhaustivas* y *mutuamente excluyentes*. La *exhaustividad* alude al hecho de que deben existir suficientes categorías para clasificar todas nuestras observaciones. Por ejemplo, la clasificación de la variable estado civil en las categorías casado, soltero, divorciado y viudo no cumple el requisito de la exhaustividad porque no permite clasificar uniones homosexuales y tampoco las parejas heterosexuales que no están legalmente casadas.

Mutuamente excluyente refiere a la necesidad de clasificar cada una de las observaciones en una y solo una categoría. Por ejemplo, si nosotros trabajamos con la variable tipos de revistas leídas, debemos definir la variable de tal manera de que cada una de nuestras observaciones puedan ser clasificadas en una y sólo una de sus categorías. Si el sistema de categorías es: deportes, femeninas, actualidad, sensacionalistas, humor y actualidad política; las categorías actualidad y actualidad político no son mutuamente excluyentes porque una revista de actualidad política puede ser considerada una revista de actualidad a secas, y entonces, sus lectores podrían ser clasificados en ambas categorías.

Las unidades de análisis

En la investigación social, los investigadores pueden elegir qué o a quiénes estudiar; es decir, eligen **aquellas unidades de análisis sobre las que van a construir descripciones y que van a comparar para describir y/o explicar las diferencias. Las unidades de análisis son las personas, grupos, organizaciones, productos sociales, etc. cuyas características observan, describen y explican los investigadores sociales.**

En general, los científicos sociales escogen *individuos* como sus unidades de análisis. Se observan sus características (sexo, edad, tipo de revista que lee, etc.) y se reúnen para dar una imagen del grupo al que pertenecen los individuos. Por ejemplo, se puede registrar la edad y el género de los estudiantes que en el primer semestre del 2005 se encuentran cursando Metodología y Técnicas de la Investigación Social y caracterizar al grupo diciendo que el 53% de los estudiantes son varones y el 47% mujeres, con una edad promedio de 19.5 años. Aunque la descripción final sería del conjunto de la clase, las características individuales se reúnen para caracterizar al grupo mayor.

Este es el caso de la investigación de Landi et al (1990); sus unidades de análisis son hombres y mujeres que residen en la Ciudad de Buenos Aires y en el Conurbano Bonaerense. Los autores observan sus características y las reúnen para construir *perfiles de consumos culturales* (caracterizar a los que miran TV, los que consumen medios gráficos, etc.).

Las organizaciones o grupos sociales también suelen ser unidades de análisis en la investigación científica. Entre los ejemplos se cuentan las escuelas, las universidades, los supermercados, las familias, las empresas que producen bienes y/o servicios, etc. Cada escuela se podría caracterizar en términos de la cantidad de alumnos, el número de maestros que trabajan en el establecimiento, la cantidad de becas comedor que reciben del gobierno, el % de su matrícula que son extranjeros, etc.

En el marco de una investigación *es absolutamente esencial determinar cuáles son las unidades de análisis para poder decidir qué observaciones hacer sobre qué y quiénes.*

Tabla 2
Ejemplos de Unidades de análisis

<u>Individuos como unidades de análisis</u>	<u>Familias como unidades de análisis</u>
¿Cuántos años tienes?	¿Cuántos niños menores de 6 años hay en la familia?
¿Estás afiliado a algún partido político?	¿Cuál es el ingreso total familiar?
¿Cuál es tu profesión?	¿Cuál es el promedio de años de escolaridad de los miembros adultos del hogar?
<u>Organizaciones como unidades de análisis</u>	<u>Ciudades como unidades de análisis</u>
¿Cuántos empleados tiene la empresa?	¿Cuál fue la tasa de homicidios del año 2004?
¿Cómo es la composición de género (% de varones y de mujeres)?	¿Cuál es la densidad de población?
¿Qué tipo de bienes y/o servicios produce?	¿Cuál es la superficie?

Proceso de operacionalización

Recapitulando...

Si tomamos con referencia la investigación de Landi et al (1990), hasta aquí hemos identificado:

- (1) Pregunta de investigación que alude a *¿cuál es el perfil de los consumos culturales de la Ciudad de Buenos Aires y del GBA? Expresada de otro modo ¿cuáles son los hábitos/ comportamientos y gustos de porteños y bonaerenses que permiten identificar públicos diferenciales de las ofertas comunicativa y cultural?*
- (2) Unidad de análisis: hombres y mujeres de entre 15 y 64 años de edad, que residen en la Ciudad de Buenos Aires y en el Conurbano Bonaerense.
- (3) Hipótesis que plantea relaciones entre variables: los públicos y los consumos culturales se encuentran diferenciados por razones económicas, de competencia educativa de los entrevistados y, también, por una diversidad regida por distintos gustos, tradiciones, gramática de desciframiento e identidades.
- (4) Variables: tipos de consumos culturales, tipo de oferta comunicativa, nivel socio económico, nivel de instrucción, lugar de residencia.

Ahora bien, analicemos con un poco más de detenidamente los conceptos involucrados en la hipótesis planteada por Landi et al (1990) ¿Cuáles son los conceptos involucrados en la hipótesis? *Públicos, consumos culturales, razones económicas, competencias educativas*, entre otros. Estos conceptos son *conceptos que están en el dominio de la teoría*. Es decir, son conceptos que no pueden observarse directamente o que no es posible poner en correspondencia con lo empírico (Cortés y Rubalcava, 1997).

Sin embargo, para avanzar en el desarrollo de la investigación fue necesario establecer las vinculaciones que existen entre estos conceptos y lo real. Los conceptos teóricos se ligan con lo real a través de *conceptos observables*; estos conceptos son utilizados por los investigadores como *mediaciones* entre el cuerpo teórico y lo real construido.

¿Cuáles son estos conceptos en el marco de la investigación de Landi et al (1990)?

Tabla 3

Ejemplos de relaciones entre conceptos teóricos y conceptos observables

Conceptos teóricos	Conceptos observables
Consumos culturales	Prácticas culturales de los agentes
Razones económicas	Nivel socio económico
Competencias educativas	Nivel de instrucción

A pesar de que los conceptos observables son más tangibles que los conceptos teóricos aún resulta difícil observarlos en la base empírica. Para poder realizar esta operación de observar/ contrastar los conceptos observables con la base empírica es necesario diseñar *indicadores* que nos permitan recortar un segmento de lo real y en esa interacción entre observación y realidad construir un dato. *Los indicadores son el producto de la mirada conceptual sobre la realidad* (Cortés y Rubalcava, 1997:17). Lo que caracteriza a todo indicador es *que existe en el mundo empírico y que son susceptibles de registrarse. El dato es el resultado del registro empírico de los indicadores.*

Hasta ahora hemos supuesto implícitamente que a cada concepto le corresponde un indicador, pero si aplicamos esa regla a los conceptos observables que propone Landi et al (1990) en su investigación creo que rápidamente advertiremos que eso no muy posible. Para observar (ergo medir) esos conceptos es necesario contar con una *batería de indicadores* que me permitan por ejemplo, dar cuenta de manera cabal del nivel socio económico de los entrevistados o de sus prácticas culturales.

Este proceso de transformación de conceptos teóricos en conceptos observables y luego en indicadores es lo que se denomina proceso de operacionalización.

Cortés y Rubalcava (1997:16) grafican este proceso con un esquema que les propongo que revisen.

Una idea general que se desprende del gráfico y que me interesa enfatizar es que *los indicadores sólo tienen sentido si están relacionados con un cuerpo teórico. Los indicadores son los referentes empíricos de concepto observables que reflejan propiedades observables de las unidades de análisis.* Así entendidos los indicadores, sólo tienen pleno sentido cuando se llenan de contenido conceptual.

Volvamos entonces a uno de nuestros conceptos para poder dar cuenta de qué indicadores fueron los que se utilizaron para su medición. Retomemos entonces el concepto de *consumos culturales*; dijimos que este concepto se corresponde con el concepto observable *prácticas culturales*.

Sin embargo, este concepto *prácticas culturales* es aún un concepto abstracto, difícil de contrastar empíricamente. Es necesario que definamos los referentes empíricos (indicadores) de este concepto de manera tal que sea posible clasificar a nuestras unidades de análisis. En el marco de la investigación de Landi et al (1990), el concepto *prácticas culturales* se tradujo a nivel operativo como *hábitos de consumo, preferencias de consumo, usos y gustos* en relación a tres grandes tipos de consumos (TV, lecturas y radio y música).

Veamos algunos ejemplos sobre qué indicadores se han elegido para medir *hábitos de consumo, preferencias de consumo, usos y gustos*.

En relación a la TV, los hábitos de consumo, las preferencias, los usos y los gustos fueron indagados a partir del:

- Tipo de programa que suele ver el televidente (hábitos de consumo)
- Razones acerca de por qué ve televisión (gustos)
- Tipo de programas que le gusta ver (preferencias de consumo)
- Canales de TV que ve habitualmente (hábitos de consumo)
- Horas a la semana que ve TV (hábitos de consumo)
- Tipo de programas eliminaría de la TV (gustos)
- Tipo de programas agregaría en la TV (gustos)
- Lugar privilegiado para mirar TV (hábitos de consumo)

En relación a las lecturas, se midieron a través de:

- Lectura de algún diario en la última semana (hábitos de consumo)
- Frecuencia de lectura de los diarios (hábitos de consumo)
- Diario o diarios que lee regularmente (hábitos de consumo)
- Lectura de los suplementos de los diarios
- Tipo de suplementos que lee (hábitos de consumo)

- Compra/ adquisición de fascículo (hábitos de consumo)
- Razones de comprar/ adquisición este tipo de lecturas (usos)
- Lectura de revistas (hábitos de consumo)
- Frecuencia de lectura de las revistas (hábitos de consumo)
- Tipo de revistas que le gusta leer (preferencias de consumo)

Tal como podemos observar, lo que el investigador intenta a través del *proceso de operacionalización* es reducir, constructor, conceptos, ideas o hipótesis a datos. En este proceso entran de lleno las *técnicas de recolección de datos*. Es decir, instrumentos y procedimientos que le permiten al investigador transformar lo real construido en datos significativos en el marco del proceso de investigación.

En el marco de las estrategias de investigación cuantitativas, la técnica de uso más frecuente es *la encuesta*. Esta técnica es la que utilizan Landi et al (1990) para recoger la información y construir los datos que finalmente les permiten caracterizar/ analizar los públicos y consumos culturales de la Ciudad de Buenos Aires y del Conurbano.

La *encuesta* es una técnica que utiliza un conjunto de procedimientos estandarizados de investigación mediante los cuáles se recogen y analizan datos, en general *obtenidos con base en una muestra de unidades de análisis*, sobre las que se pretende, explorar, describir y/o explicar una serie de características.

Hay distintos tipos de encuestas, encuestas personales, encuestas por correo y encuestas telefónicas, entre otros. En el caso de Landi et al. (1990), se utiliza una *encuesta personal o cara a cara*, que implica la participación de encuestadores que aplican un cuestionario a las personas involucradas en la investigación (unidades de análisis).²

Para poder avanzar en el diseño de una *encuesta* es imprescindible que el investigador haya definido adecuadamente su pregunta de investigación (es decir su problema de investigación) y haya avanzado en el proceso de traducción de sus conceptos teóricos en indicadores. Asimismo, será necesario tener en claro cuáles son las variables que actúan como factores determinantes o estructurantes del fenómeno a investigar.

Por ejemplo, volvamos al caso de Landi et al (1990), tal como vimos anteriormente, el fenómeno que los autores pretender abordar es el de los *públicos y consumos culturales*. Para poder avanzar en la indagación del fenómeno vimos como ellos construyen las diferentes mediaciones que les permiten vincular el concepto *consumos culturales* con el mundo empírico. Pero para poder avanzar en la definición y caracterización de los públicos, además de conocer los consumos culturales³ fue necesario identificar las variables que permiten al investigador encontrar diferentes *perfiles de consumo*. En el marco de la investigación, esas variables son *nivel socio económico, nivel de instrucción, lugar de residencia, sexo, edad*, entre otras.

La relación entre los distintos conceptos estaba planteada en las hipótesis de la investigación, pero es imprescindible tener claridad sobre ellos en el momento de

² Interesa destacar que si bien las personas son las unidades de análisis más frecuentes de las encuestas; esta técnica también puede ser utilizada para realizar investigaciones sobre otras unidades de análisis.

³ Definidos a nivel operacional como *hábitos de consumo, preferencias de consumo, usos y gustos*.

organizar la encuesta porque lo que el investigador deje de medir en la encuesta ya no podrá ser utilizado como dato.

El instrumento que se utiliza para recoger los datos en el marco de una encuesta es el *cuestionario*. El proceso de diseño del cuestionario se inicia con la traducción de los indicadores en preguntas tal y como se formulan a los entrevistados. Se trata de la traducción del fenómeno a investigar en preguntas organizadas en un cuestionario. Los conceptos y sus indicadores son incorporadas al cuestionario en preguntas que se formulan a cada encuestado.

Este proceso de traducción es parte esencial de la técnica de encuesta. La mayoría de los temas que se revelan mediante encuestas no pueden ser pasados directamente a los entrevistados en forma de preguntas. El mal uso y el fracaso de las encuestas, en general, se debe a la ingenua equiparación entre el fenómeno a investigar y las preguntas del cuestionario.

Seguidamente, a la necesidad de traducir los indicadores a preguntar, se suma la de formular dichas preguntas. Es decir, una vez que fijamos el contenido de la pregunta hay que realizar un trabajo de formulación.

Tabla 4
Ejemplos de indicadores su correspondencia con preguntas de cuestionario

Indicador	Pregunta del cuestionario
TV	
Tipo de programa que suele ver el televidente (hábitos de consumo)	¿Qué tipo de programas suele ver usted? (hábitos de consumo)
Razones acerca de por qué ve televisión (gustos)	¿Por qué ve televisión? (gustos)
Tipo de programas que le gusta ver (preferencias de consumo)	¿Qué tipo de programas son los que le gusta ver? (preferencias de consumo)
Canales de TV que ve habitualmente (hábitos de consumo)	¿Qué canales de TV son los que ve habitualmente? (hábitos de consumo)
Horas a la semana que ve TV (hábitos de consumo)	¿Cuántas horas a la semana ve de TV? (hábitos de consumo)
Tipo de programas eliminaría de la TV (gustos)	¿Qué tipo de programas eliminaría de la TV? (gustos)
Tipo de programas agregaría en la TV (gustos)	¿Qué tipo de programas agregaría en la TV? (gustos)
Lugar privilegiado para mirar TV (hábitos de consumo)	¿Cuál es su lugar privilegiado para mirar TV? (hábitos de consumo)
Lecturas	
Lectura de algún diario en la última semana (hábitos de consumo)	¿Ha leído Usted algún diario en la última semana? (hábitos de consumo)
Frecuencia de lectura de los diarios (hábitos de consumo)	¿Con qué frecuencia lee Usted los diarios? (hábitos de consumo)
Diario o diarios que lee regularmente (hábitos de consumo)	¿Qué diario o diarios lee Usted más regularmente? (hábitos de consumo)
Lectura de los suplementos de los diarios	¿Lee usted los suplementos de los diarios (hábitos de consumo)
Tipo de suplementos que lee (hábitos de consumo)	¿Cuáles? (hábitos de consumo)
Compra/ adquisición de fascículo (hábitos de consumo)	¿Ha comprado Usted o su familia en el último año algún tipo de fascículo? (hábitos de consumo)
Razones de comprar/ adquisición este tipo de lecturas (usos)	¿Por qué razones decidió comprar/ adquirir este tipo de lecturas? (usos)
Lectura de revistas (hábitos de consumo)	¿Ha leído Usted alguna revista? (hábitos de consumo)

Frecuencia de lectura de las revistas (hábitos de consumo)	¿Con qué frecuencia lee Usted las revistas? (hábitos de consumo)
Tipo de revistas que le gusta leer (preferencias de consumo)	¿Qué tipo de revistas son las que le gusta leer? (preferencias de consumo)

Fuente: Landi et al, 1990

Este proceso de traducción y de formulación debe ser cuidadosamente controlado a la luz de la teoría y la experiencia metodológica pues a lo largo del proceso la conexión entre los objetivos de la investigación y el cuestionario se hace cada vez más difusa.

Una vez que el investigador ha transitado por todas estas fases del proceso de investigación; es decir, una vez que ya tiene claro cuál es el fenómeno que desea investigar, qué conceptos va a usar, cómo esos conceptos se vinculan con el mundo empírico. Incluso una vez que efectivamente ha diseñado su cuestionario, el investigador está en condiciones de avanzar en la realización del *trabajo de campo*.⁴ El trabajo de campo es el momento/ etapa de la investigación en la que el investigador procede a recoger los datos, en nuestro caso a través de la aplicación de una encuesta. A través de la aplicación de la encuesta el investigador puede conocer el valor que las variables toman en cada unidad de análisis; es decir, puede conocer que Juan tiene 33 años y lee diariamente Página 12 y que además ve en promedio 2 horas de TV.

Ahora bien ¿cómo hicieron Landi et al (1990) para pasar de la información recogida con base en la aplicación de su encuesta a la presentación de datos significativos sobre públicos y consumos culturales?

La matriz de datos

El primer paso en este proceso de *procesamiento de la información* es organizar unidades de análisis, indicadores relevados y el valor que asume el indicador para cada unidad de análisis en una *matriz de datos*. La matriz de datos es un arreglo rectangular con tantos renglones como unidades de análisis haya y con una columna por indicador. Las casillas de la matriz definidas por la intersección de renglones y columnas contienen los valores de los indicadores (Cortés y Rubalcava, 1997:232)

De este modo, la matriz de datos permite ordenar la información de manera tal que es posible visualizar los tres elementos que componen el dato (Abritta, 1999): las unidades de análisis, las variables y los valores o categorías de las variables.

⁴ Previamente a la realización del trabajo de campo, el investigador deberá establecer y diseñar el tipo de muestra a la que se aplicará la encuesta y proceder a la selección de las unidades de análisis que formarán parte de la misma. Este tema será desarrollado en el Teórico 6/ Módulo 2.

Tabla 5
Ejemplo de matriz de datos

Unidades de análisis	Variable 1 Nombre	Variable 2 Edad	Variable 3 Lugar de Residencia	Variable 4 Tipo de programa que suele ver	Variable 5 Canales de TV que ve habitualmente	Variable 6 Horas a la semana que ve TV	Variable n
Unidad 1	Juan	22	CBA	Entretenimiento/ juegos	13	2	
Unidad 2	Romina	15	CBA	Entretenimiento/ juegos	2	4	
Unidad 3	José	16	GBA	Entretenimiento/ juegos	11	3	
Unidad 4	Ana	35	GBA	Noticieros	13	2	
Unidad 5	Clara	27	CBA	Películas	11	5	
Unidad 6	Rodolfo	46	GBA	Humor y comedias	9	1	
Unidad 7	Josefina	24	CBA	Películas	11	4	
Unidad 8	Pedro	18	CBA	Novelas	11	4	
Unidad n							

La construcción de la matriz de datos es fundamental porque su construcción da cuenta de las complejas operaciones teórico/ metodológicas involucradas en todo proceso de investigación, al mismo tiempo que es una herramienta indispensable para proceder al análisis de los datos.

Distribuciones de frecuencias

Y ¿que sigue después de esto...?

En una investigación el investigador, en general, tiene que trabajar con una gran cantidad de datos. Por ejemplo, en su investigación Landi et al (1990) recogieron información sobre un gran número de variables con base en una muestra de 600 unidades de análisis (su matriz tenía 600 renglones). Para darle sentido a esta masa de información, los investigadores debemos organizar y resumir esa información de manera sistemática. El método más básico para organizar datos es clasificar nuestras observaciones en una *distribución de frecuencias*. Una *distribución de frecuencias* es una tabla que permite visualizar rápidamente el número de observaciones que cae en cada una de las categorías de la variable que estamos analizando.

La organización de los datos en *distribuciones de frecuencias* permite realizar un análisis exhaustivo de cada una de las variables incluidas en la matriz de datos (análisis univariado).

Veamos un ejemplo con base en la investigación de Landi et al (1990)...

Tabla 6**Medio que desde la perspectiva de los entrevistados es el más creíble. CBA y GBA.**

Medio que desde la perspectiva de los entrevistados es el más creíble	Frecuencia (f)
Ninguno	176
Diarios	161
Radios	126
Televisión	107
Revistas	7
Nr/Ns	23
Total (n)	600

Fuente: Landi et al, 1990:69.

Nótese que la distribución de frecuencias es organizada en una tabla, que tiene un número (Tabla 6) y un título descriptivo. El título indica el tipo de datos que se presenta: ¿Cuál es el medio al que los entrevistados le creen más? La tabla consiste en dos columnas. La primera columna identifica a la variable (Medio que desde la perspectiva de los entrevistados es el más creíble) y sus categorías (Ninguno, Diarios, Radios, Televisión, Revistas, Nr/Ns). La segunda columna titulada “Frecuencia (f)”, indica el número de unidades de análisis en cada categoría de la variable así como el total de casos que integraron la muestra (n = 600). Nótese también que la fuente de la cual se extrajeron los datos está claramente especificada en la base de la tabla.

¿Qué podemos aprender de la información consignada en la Tabla 6? La tabla muestra que, en 1990, de entre 600 personas residentes en la Ciudad de Buenos Aires y en el Conurbano Bonaerense, 176 consideraba que ningún medio de comunicación era creíble. Fuera de este grupo, 161 personas consideraron a los diarios como los medios más creíbles, seguidos por las radios y la televisión. El medio de menor credibilidad son las revistas.

Las distribuciones de frecuencias son útiles para presentar información de manera compacta. Si embargo, cuando el número de casos es muy grande, como por ejemplo ocurre cuando se procesan resultados de censos de población, las frecuencias resultan difíciles de comprender. Para facilitar el entendimiento y estandarizar las frecuencias, es conveniente transformarlas en frecuencias relativas - es decir, en *porcentajes*.

Tabla 7**Medio que desde la perspectiva de los entrevistados es el más creíble. CBA y GBA.****En %.**

Medio que desde la perspectiva de los entrevistados es el más creíble	f%
Ninguno	29,3
Diarios	26,8
Radios	21,0
Televisión	17,8
Revistas	1,2
Nr/Ns	3,8
Total (n)	100%

Fuente: Landi et al, 1990:69.

Un porcentaje es una frecuencia relativa obtenida a partir de dividir la frecuencia que corresponde a cada categoría de la variable por el número total de casos de la muestra (n) y multiplicarla por 100.

$$\text{Porcentaje (\%)} = f / n * 100$$

En donde:

f = frecuencia

n = número total de casos de la muestra

Análisis univariado

Para analizar exhaustivamente las variables incluidas en la matriz de datos, también, es posible emplear gráficos y estadísticos que permiten estimar, de manera precisa, características de una población con base en datos de la muestra. El uso de los gráficos y de los estadísticos *depende básicamente del tipo de variables con las que el investigador trabaje y de su nivel de medición*. Los estadísticos univariados se agrupan en tres grandes grupos:

- a. Medidas de tendencia central
- b. Medidas de dispersión
- c. Medidas de forma de la distribución

Nosotros nos concentraremos en las medidas de tendencia central y de dispersión.

Veamos esto con un poco más de detalle...

El análisis estadístico de los datos en el marco de una investigación habitualmente involucra operaciones matemáticas, que van desde el simple conteo hasta operaciones de mayor complejidad. Sin embargo, no todas las operaciones pueden utilizarse con todas las variables. El tipo de operaciones estadísticas depende de cómo nuestras variables están medidas.

Existen diferentes tipos de escalas de medición que se distinguen de acuerdo a la rigurosidad con que han sido construidas y al propio comportamiento de las variables que miden. Se acostumbra a clasificarlas en cuatro tipos generales que son los siguientes: *escalas nominales, ordinales, de intervalos y de cocientes o razones*.

Escalas nominales son aquéllas en que sólo se establece una equivalencia entre la escala y los diferentes valores que asume la variable. Es como una lista de las diferentes posiciones que pueda adoptar la variable, pero sin que en ella se defina ningún tipo de orden o de relación. En la investigación de Landi et al (1990), los investigadores consideraron importante clasificar a sus entrevistados según el género para observar si existen diferencias entre los consumos culturales de mujeres y hombres. Los distintos valores que reconoce la variable género son varón y mujer. Entre estos valores no cabe obviamente ninguna jerarquía, no se puede trazar ningún ordenamiento. Sin embargo, a la enunciación explícita de estas posibilidades la consideramos como una escala, pues de algún modo es útil para medir el comportamiento de la variable, indicándonos en qué posición se halla cada unidad de análisis.

Las *escalas ordinales* distinguen los diferentes valores de la variable jerarquizándolos simplemente de acuerdo a un rango. Establecen que existe una gradación entre uno y otro valor de la escala, de tal modo que cualquiera de ellos es mayor que el precedente y menor que el que le sigue a continuación. Sin embargo, la distancia entre un valor y otro queda indeterminada. En otras palabras, tales escalas nos esclarecen solamente el rango que las distintas posiciones guardan entre sí. Un ejemplo de escala ordinal es el nivel de instrucción: podemos decir que una persona que ha completado la escolaridad primaria de instrucción ha recibido más instrucción que quien no la completó y menos que quien posee un título universitario. Sin embargo no puede afirmarse válidamente que la diferencia entre quien posee primaria completa y quien no a completado ese nivel es igual a la diferencia que existe entre quienes han interrumpido los estudios universitarios y quiénes lograron

completarlos. Por tanto, como no podemos determinar la equivalencia entre las distancias que separan un valor de otro, debemos concluir que la escala pertenece a la categoría ordinal.

Las *escalas de intervalos*, además de poseer la equivalencia de categorías y el ordenamiento interno entre ellas, como en el caso de las ordinales, tienen la característica de que la distancia entre sus intervalos está claramente determinada y que estos son iguales entre sí. Un ejemplo típico de las escalas de intervalos iguales está dado por las escalas termométricas. Entre 23 y 24 grados centígrados, por ejemplo, existe la misma diferencia que hay entre 45 y 46 grados. Muchas otras escalas, como las que se utilizan en los test psicológicos y de rendimiento, pertenecen a este tipo. La limitación que poseen es que no definen un cero absoluto, un valor límite que exprese realmente la ausencia completa de la cualidad medida. Por ello no se pueden establecer equivalencias matemáticas como las de proporcionalidad: no puede afirmarse que 24°C es el doble de temperatura que 12°C, porque el cero de la escala es un valor arbitrario y no se corresponde con la ausencia absoluta de la variable que se mide.

Por último tenemos las *escalas de cocientes*, llamadas también de razones. En ellas se conservan todas las propiedades de los casos anteriores pero además se añade la existencia de un valor cero real, con lo que se hacen posibles ciertas operaciones matemáticas, tales como la obtención de proporciones y cocientes. Esto quiere decir que un valor de 20 en una escala de este tipo es el doble de un valor de 10, o de las dos terceras partes de un valor de 30. Son escalas de cocientes las que miden la longitud, la masa, la intensidad de corriente eléctrica y otras variables del mundo físico. Difícilmente las variables que intervienen en las ciencias sociales son medidas con escalas de razones, pues son contados los casos en que dichas variables pueden ser definidas con la exactitud y precisión necesarias. La economía y la demografía son, entre estas disciplinas, las que más utilizan escalas de razones.

Medidas de tendencia central

Describen cómo se agrupan los valores de una *variable* alrededor de un “valor típico” de la distribución. Por lo que, proporcionan una síntesis de la información contenida en la distribución.

Las medidas de *tendencia central* más empleadas en la investigación social son las siguientes: *la media*, *la mediana* y *la moda*.

- 1) La *media* es la medida más representativa, siempre y cuando la variable sea cuantitativa (de *intervalo* o de *razón*). Su cálculo precisa de la participación de todos los valores de la distribución. Cada uno de ellos se multiplica por sus respectivas *frecuencias absolutas*. Después, se suman todos los productos, y el resultado se divide por el número total de casos. De esta forma se obtiene el *promedio* de los valores de la distribución, que es como se define la *media aritmética*. El *inconveniente* fundamental de este estadístico es que se ve afectado por *valores* muy extremos en una distribución. Cuando la variable incluye valores muy dispares (o extremos), la *media* pierde representatividad a favor de la *mediana*, como *medida de tendencia central*.
- 2) La *mediana* es el valor que divide a la distribución en dos partes iguales. Se sitúa en el medio de la distribución. Lo que permite el conocimiento de los *valores* de mayor representación en la *muestra*. Su cálculo precisa que la variable sea, al menos, ordinal.
- 3) La *moda* denota el valor de mayor frecuencia de una distribución; aquel que más casos comparten. La distribución puede ser *unimodal* (una sola *moda*),

pero también *bimodal* (dos *modas*), o *multimodal* (más de dos *modas*). Esto dificulta su interpretación, en caso de coexistir varias *modas*. A este *inconveniente* primordial se suma otro importante: en su cálculo no intervienen todos los *valores* de la distribución. Si bien, su principal *ventaja* es su universalidad. Puede estimarse para cualquier tipo de variable, ya que el nivel de medición mínimo exigido es el *nominal*.

Además de estos estadísticos, existen los *cuantiles* como *medidas de tendencia no central*. Estos representan *valores* que dividen a la distribución en partes iguales. Los *cuartiles* la dividen en cuatro partes iguales (cada una de ellas incluye al 25% de los valores); los *deciles* en diez partes; y los *percentiles* en cien partes. Su cálculo se asemeja al de la *mediana*. De hecho, el *cuartil* dos, el *decil* cinco, y el *percentil* cincuenta, expresan el *valor mediano*.

Medidas de dispersión

Al conocimiento de los *valores centrales* de la distribución le sigue la medición de su representatividad: la mayor o la menor *variabilidad* existente en torno a la *media* o a la *mediana* de la distribución.

Las medidas de *dispersión absolutas* más comunes son el *rango* (o recorrido), la *desviación típica*, y la *varianza*.

- 1) El *rango* o *recorrido* expresa el número de valores incluidos en la distribución. Estos se obtienen de la diferencia entre el valor superior y el inferior. Su comprensión es sencilla, aunque presenta, en su contra, un *inconveniente* importante: es susceptible a la distorsión proporcionada por valores muy extremos en una distribución. Ello se debe a la única consideración del primer y último valor de la distribución.

Para solventar esta deficiencia, suelen aplicarse otras medidas de rango, que emplean un mayor volumen de información. Si bien, éstas exigen que la variable sea, al menos, ordinal. Se trata de los recorridos: intercuartílico (la diferencia entre el tercer cuartil y el primero; por lo que incluye al 50% de los valores centrales de la distribución), semiintercuartílico (el 25% de los valores centrales), interpercentílico (el 80% de los valores centrales, al ser la diferencia entre el percentil 90 y el 10), y el semiinterpercentílico (el 40%).

- 2) La *desviación típica* es el promedio de la desviación de los casos con respecto a la *media*. Como indicador de heterogeneidad (o de dispersión de los valores de una distribución), su estimación se exige siempre que se calcule la media, porque ayudará a la interpretación de su representatividad en la distribución.

Al igual que la *media*, su valor viene expresado en la unidad de medición de la *variable*, y únicamente puede calcularse cuando la variable es *cuantitativa*.

- 3) La *varianza* constituye otra medida de heterogeneidad de una distribución. Se define como el cuadrado de la *desviación típica*. Su valor expresa el grado de heterogeneidad de una población respecto a la *variable* medida, siendo sus características similares a la *desviación*.

A partir de la *desviación*, puede calcularse una medida de *dispersión relativa* que ayudará en su interpretación. Se trata del *coeficiente de variabilidad de Pearson*. Una medida estandarizada que se obtiene del cociente entre el valor de la *desviación* y la *media aritmética*. Su valor se expresa en porcentajes, siendo de utilidad en la comparación de la homogeneidad de dos o más grupos respecto a una o más variables.

Por último, cuando se calcula la *mediana*, también suelen estimarse estadísticos que midan su representatividad en la distribución. Concretamente, la *desviación media de la mediana* y el *coeficiente de variabilidad de la mediana de Pearson* (obtenido del cociente entre la *desviación media de la mediana* y la *mediana*). Su interpretación es análoga a los estadísticos anteriores.

Análisis bivariado

Hasta aquí trabajamos con bastante detenimiento algunas de las técnicas de análisis univariado: trabajamos con distribuciones de frecuencias (absolutas y relativas) y, también, hicimos lo propio con las medidas de tendencia central. Todas estas técnicas se refieren sólo a una variable, constituyen elementos de estadística univariada.

En el caso de la investigación de Landi et al (1990), nos detuvimos en el análisis de la distribución de la variable *género*, de la variable *edad*, de la variable *medio que desde la perspectiva de los entrevistados es el más creíble*.

Sin embargo, en la mayoría de los casos, a los investigadores no nos interesa analizar variables sueltas. Nos interesa detectar y describir la existencia de relaciones entre variables. En realidad, la existencia de relación entre los *conceptos teóricos* operacionalizados a través de esas variables.

Lo que el investigador busca es analizar si existe o no relación entre los conceptos que están en juego en sus hipótesis.

Relaciones entre dos variables

El análisis bivariado es un procedimiento destinado a detectar y a describir la relación entre dos variables. Hasta aquí, ustedes tenían una aproximación intuitiva al término *relación*. En este encuentro vamos a analizar el concepto de *relación entre variables* con mayor profundidad.

En parte esta aproximación intuitiva se funda también en que los medios de comunicación todo el tiempo nos informan acerca de relaciones entre fenómenos y/o procesos. Por ejemplo, uno de los artículos del diario Clarín que ustedes trabajaron como material de prácticos, planteaba que la TV violenta influye sobre el comportamiento de los chicos.⁵ El artículo presenta los resultados de un estudio que compara las conductas de niños con antecedentes penales y sin antecedentes y muestra que los primeros sentían mayor grado de identificación, que los segundos, con personajes de películas o series con alto contenido de violencia. Este sencillo ejemplo ilustra una *relación* o *asociación* entre la variable *conducta delictiva*⁶ y la variable *grado de identificación con personajes violentos*.

La búsqueda de relaciones entre variables está en el corazón de todo proceso de producción de conocimiento. El investigador, en general, se hace preguntas que ponen en relación fenómenos como, por ejemplo, mirar películas y series con personajes violentos y la existencia de conductas delictivas. En el caso de Landi et al (1990), la pregunta refiere a cómo se vinculan los hábitos/ comportamientos y gustos de porteños y bonaerenses con su nivel socioeconómico o con sus capacidades educativas.

⁵ Clarín, 20 de Febrero, 2005.

⁶ Cuyas categorías serían *si* y *no*; es decir presencia o ausencia del atributo conducta delictiva.

En cada uno de estos ejemplos, *relación* significa que ciertos valores de la variable tienden “a ir junto” con ciertos valores de la otra variable. En el primer ejemplo la presencia de conductas violentas tiende “a ir junto” con un alto grado de identificación con personajes de películas o series con alto contenido de violencia. En el caso de Landi et al (1990: 74), los autores llegan a la conclusión de que los sectores con más educación y nivel socioeconómico son los que creen más en los medios.

Al plantear la existencia de relación entre dos variables estamos postulando que ambas variables *covarian*; es decir, que parte de la varianza de la variable dependiente se explica por la varianza de la variable independiente. Usamos el concepto de *covarianza* para expresar en qué medida al moverse una variable se mueve la otra: en qué medida al analizar las preferencias de personas de diferente nivel socioeconómico varía sus pautas en los consumos culturales.

Vamos a introducirnos, entonces, en una técnica que nos permitirá analizar las relaciones entre dos variables. Esta técnica es la construcción de *cuadros bivariados o de contingencia*.

Sin embargo, antes debemos asentar algunas nociones previas...

Variable independiente y variable dependiente

En la investigación en ciencias sociales, las hipótesis habitualmente se enuncian en términos de una relación entre una *variable independiente* y *variable dependiente*. Identificar con claridad la diferencia entre *variable independiente* y *variable dependiente* es importante si el objetivo es proveer una explicación (aunque sea provisoria y parcial) sobre un fenómeno social.

Por ejemplo, en la hipótesis que postula Landi et al (1990) queda claro que *los públicos y los consumos culturales se encuentran diferenciados por razones económicas, de competencia educativa de los entrevistados y, también, por una diversidad regida por distintos gustos, tradiciones, gramática de desciframiento e identidades*. Tal como lo postulan Landi et al (1990), las diferencias entre los públicos y entre los consumos culturales se explican por otros factores, que en este caso son, fundamentalmente, el nivel socioeconómico y el nivel de instrucción.

En el lenguaje de la investigación, la variable que el investigador quiere explicar es denominada *variable dependiente*. La variable que se espera “explique” o de cuenta de las variaciones en la variable dependiente se denomina *variable independiente*. En el caso de la investigación que nos ocupa *las pautas de los consumos culturales y los perfiles de los consumidores* constituyen las *variables dependientes*. Mientras que, el *nivel socioeconómico y el nivel de instrucción*, entre otras, son *variables independientes* ¿Qué quiere decir esto? Sencillamente que *las pautas de los consumos culturales y los perfiles de los consumidores varían según su nivel socioeconómico y su nivel de instrucción*.

Me gustaría dejar en claro que el hecho de postular variables dependientes y variables independientes en el contexto de una investigación no necesariamente equivale a postular una relación causal entre ellas. Utilizamos esta denominación aún cuando la relación entre las variables no está necesariamente articulada en términos de *causa y efecto* ¿Por qué digo esto? Porque en ciencias sociales -- por las características de su objeto de estudio -- no es fácil inferir relaciones de causalidad entre variables.

¿Cómo identificar cuándo una variable se postula como dependiente y cuándo como independiente? Veamos algunas ayudas que nos pueden ayudar a identificar cuál es el rol de las variables en las hipótesis que plantean los investigadores.

1. La variable dependiente es siempre el fenómeno que el investigador intenta explicar, estudiar, investigar, describir y/o analizar. Es siempre el objeto de la investigación.
2. La variable independiente, en general, es temporalmente anterior a la variable dependiente.
3. La variable independiente es vista como un factor que directa o indirectamente influencia el comportamiento de la variable dependiente.

En el mundo real, las variables no son ni dependientes ni independientes, su rol en la investigación depende de cómo el investigador defina su problema de investigación. De este modo, una variable que en una investigación asume el rol de variable dependiente, puede ser independiente en el contexto de otra investigación.

¿Cómo construir un cuadro de contingencia o cuadro bivariado?

Un cuadro bivariado despliega la distribución de una variable según las categorías de otra variable. Se obtiene clasificando unidades de análisis con base en los valores que obtienen para dos variables. Puede pensarse como distribuciones de frecuencias que se relacionan/ se unen para hacer una única tabla. Para ejemplificar el proceso de construcción de un cuadro de doble entrada o cuadro bivariado vamos a utilizar como base la Tabla 12 presentada en el Anexo. La Tabla 12 representaba una muestra de encuestados clasificados según la edad, el género y el medio al cual los encuestados le creen más.

Para poder relacionar por ejemplo cómo varía la credibilidad de los medios según el género, lo primero que tenemos que hacer es construir una tabla que nos permita observar esa relación.

Tabla 8
Medio al que los entrevistados le creen más según género (en absolutos).
Submuestra.

Medio al que los entrevistados le creen más	Género		
	Varón	Mujer	
Ninguno	5	4	9
Diarios	5	3	8
TV	1	4	5
Radio	3	2	5
Revistas	1	1	2
Nr/ Nc	0	1	1
	15	15	30
	Marginales		Total de Casos (n)

Fuente: Elaboración propia con base en Landi et al, 1990.

La Tabla 8 tiene algunas características en las que quiero detenerme porque son los rasgos de típicos de la mayoría de los cuadros bivariados o de doble entrada.

1. El título de la tabla es descriptivo, identifica el contenido de la tabla en término de de las dos variables.
2. Tiene dos dimensiones, una para el género y otra para el medio al que los entrevistados le creen más. La variable *medio al que los entrevistados le creen más* está representada en las filas de la tabla, con una fila para cada una de sus categorías. La variable *género* se ubica en las columnas de la tabla, con una columna por cada situación de género identificada (varón y mujer). La tabla puede tener más columnas o más filas dependiendo de la cantidad de categorías de las variables. Habitualmente la *variable independiente* es la variable que se presenta en las columnas, mientras que la *variable dependiente* es la que se presenta en las filas.
3. La intersección entre filas y columnas se denomina *celdas*. Por ejemplo, los 5 individuos ubicados es la celda superior izquierda son varones que consideran que ningún medio es creíble.
4. Los marginales y los subtotales son las distribuciones de frecuencias de las variables *género* y *medio al que los entrevistados le creen más*, respectivamente. Las frecuencias marginales representan la distribución de frecuencias de la variable *género* y las frecuencias subtotales, la de la variable *medio al que los entrevistados le creen más*. El número total de casos que integran la muestra (n) es el número que surge de la intersección entre marginales y subtotales (todos estos elementos están indicados en la Tabla 8).
5. La fuente de la cual se extrajeron los datos que aparecen en la tabla debe estar claramente indicada en una nota en la base de la tabla.

¿Cómo calcular porcentajes en un cuadro bivariado?

Para comparar la credibilidad de los medios entre varones y mujeres, nosotros necesitamos convertir las frecuencias absolutas del cuadro en frecuencias relativas. En el caso particular de la Tabla 8, la necesidad de convertir las frecuencias absolutas en frecuencias relativas no es imprescindible porque los totales de las columnas son iguales. Sin embargo, los porcentajes son indispensables para comparar

dos o más grupos (definidos por las categorías de la variable independiente) que difieren en su tamaño (los totales de las columnas son diferentes).

Es importante que recuerden que generalmente los marginales son diferentes. Entonces, es necesario estandarizar el total de cada categoría de la variable independiente. La forma más simple de estandarizar es calculando porcentajes.

Hay dos reglas básicas para calcular y analizar %:

1. Calcular los porcentajes tomando como base la frecuencia marginal definida para cada una de las categorías de la variable independiente.
2. Interpretar la tabla comparando los porcentajes entre las categorías de la variable independiente, para cada valor de la variable dependiente.

La primera regla significa que nosotros tenemos que calcular los porcentajes tomando como base la frecuencia marginal definida para cada una de las categorías de la variable que el investigador postula como independiente. Cuando la variable independiente se consigna en las *columnas*, nosotros calculamos los porcentajes para cada una de las columnas. La frecuencia absoluta de cada celda se divide por la frecuencia marginal total de la columna a la que pertenece la celda, el total de la columna debe sumar el 100%. Cuando la variable independiente se consigna en la *fila*, los porcentajes se calculan para cada una de las filas. La frecuencia absoluta de cada celda se divide por la frecuencia marginal total de la fila a la que pertenece la celda, el total de la fila debe sumar el 100%.

En nuestro ejemplo de la Tabla 8, nosotros postulamos al género como variable independiente. Entonces nosotros usaremos el total de cada una de las categorías de la variable género (varón y mujer) como base para calcular los porcentajes. Por ejemplo, el porcentaje de varones que considera que ningún medio es creíble se obtiene dividiendo el número de varones que considera que ningún medio es creíble (5) por el total de varones en la muestra (15) y multiplicándolo por 100.

$5 / 15 * 100 = 33\%$

La Tabla 9 presenta los porcentajes calculados con base en la Tabla 8. Noten que los porcentajes de cada una de las columnas suman el 100%, incluida la columna de subtotales. Siempre mostramos los *ns* con base en los que se calculan los porcentajes - en este caso, el total de cada columna.

Tabla 9

Medio al que los entrevistados le creen más según género (en %). Submuestra.

Medio al que los entrevistados le creen más	Género		Total
	Varón	Mujer	
Ninguno	33%	27%	30%
Diarios	33%	20%	27%
TV	7%	27%	17%
Radio	20%	13%	17%
Revistas	7%	7%	7%
Nr/ Nc	0	6%	2%
Total	100%	100%	100%
	(15)	(15)	(30)

Fuente: Elaboración propia con base en Landi et al, 1990.

La segunda regla nos dice cómo comparar la credibilidad en los medios según el género de los entrevistados. La comparación debe hacerse analizando diferencias porcentuales entre las categorías de la variable independiente, para cada valor de la variable dependiente. Algunos investigadores limitan el análisis a aquellas categorías de la variable dependiente en las que se registran diferencias mayores al 10%.

Interpretando porcentajes de un cuadro bivariado

Mirando este cuadro, ¿Ustedes qué pueden decir acerca de la credibilidad en los medios?

1. Efectivamente la credibilidad en los medios es diferente entre hombres y mujeres. Es decir, hay relación entre estas dos variables ¿Por qué? ¿Cómo se advierte esto? Porque las frecuencias porcentuales se distribuyen diferencialmente según el medio de que se trate y según el género.

Si los porcentajes fuesen totalmente equivalentes entre hombres y mujeres diríamos que existe *independencia estadística* entre las variables. Igual proporción de hombres y de mujeres considera que ningún medio es creíble. El orden que le asignan a los medios según el grado de credibilidad es el mismo.

Tabla 10

Situación de independencia estadística

Medio al que los entrevistados le creen más	Género		Total
	Varón	Mujer	
Ninguno	33%	33%	33%
Diarios	33%	33%	33%
TV	7%	7%	7%
Radio	20%	20%	20%
Revistas	7%	7%	7%
Nr/ Nc	-.-	-.-	-.-
Total	100%	100%	100%
	(15)	(15)	(30)

Fuente: Elaboración propia con base en Landi et al, 1990.

2. Que para los hombres los medios son menos creíbles que para las mujeres (33% vs 27%).

3. Que los diarios son el medio en el que los hombres más confían.
4. Las mujeres, en cambio, prefieren la TV.
5. La radio tiene mayor credibilidad entre hombres que entre mujeres.
6. Hombres y mujeres juzgan de igual manera a las revistas.

Por medio de los porcentajes se hace posible la comparación ¿Qué es lo que ocurre con cada una de las categorías de la variable independiente? Es decir, que ocurre con el grupo de varones y con el grupo de mujeres ¿piensan igual? ¿Piensan diferente?

Tal como les dije anteriormente, la tabla 10 muestra la situación en donde no hay diferencias entre hombres y mujeres. Mirando la Tabla 9, en cambio, si me permite saber que para una persona que es varón existe un 33% de posibilidades de que ningún medio le resulte creíble. En vez, si es mujer, esa posibilidad se reduce al 27%.

Esta forma de analizar los cuadros bivariados nos permite saber si existe o no relación entre dos variables.

Muestreo

Hasta ahora, en esta serie de teóricos, prácticamente ignoramos la pregunta acerca de *cómo se define el quiénes o el cuáles van a ser las unidades de análisis que efectivamente van a ser relevadas* a través de, por ejemplo, la encuesta.

Los investigadores en ciencias sociales casi nunca tenemos la posibilidad de obtener información sobre el grupo completo de unidades de análisis que nos interesa investigar. A este grupo que incluye TODOS LOS CASOS (individuos, objetos, familias, instituciones, etc.) en los cuales el investigador está interesado se lo denomina *población*. Por ejemplo, en la investigación de Landi et al (1990) la población está constituida por la totalidad de las personas de 18 a 64 años que residen en la ciudad de Buenos Aires y en el Conurbano Bonaerense. A Landi y a sus colegas les fue imposible contactar a todos los habitantes de la ciudad y del Conurbano (en el año 1991, la población del Área Metropolitana de Buenos Aires ascendía a unos 11 millones de personas). Esta imposibilidad ocurre en la mayoría de los casos ya sea porque el tamaño de la población es inabarcable para cualquier mortal, porque no tenemos suficiente tiempo para relevar tantos casos y/o porque tampoco tenemos dinero suficiente para hacerlo.

Afortunadamente, tampoco es necesario relevar información de la población completa; es posible aprender mucho de esa población si nosotros seleccionamos cuidadosamente una pequeña parte de ella. A esa pequeña fracción de población la denominamos *muestra*. A través del *muestreo* - proceso de selección de una parte o fracción de la población de interés - es posible conocer características de la población con base en resultados obtenidos de un pequeño grupo (la muestra).

Esta es la base de la *inferencia estadística* - hacer predicciones o inferencias sobre la población a partir de observaciones obtenidas con base en una muestra.

En las ciencias sociales es muy común el uso de muestras para aproximarse al conocimiento de lo real. Sin embargo, para que a través de una muestra sea posible hacer inferencias sobre las características de la población es necesario que el *diseño de la muestra* y, concretamente, la selección de los casos se ajusten a determinados procedimientos.

Veamos algunos conceptos que nos van a ayudar a entender este proceso...

Parámetros y estadísticos

A las características de la población que los investigadores deseamos conocer se los denomina *parámetros*. Si continuamos con el ejemplo de Landi, dos parámetros relevantes son:

- 1- Edad media de todas las personas que leen diariamente algún periódico
- 2- El porcentaje de todos los habitantes del AMBA que tienen TV

En general, estos parámetros no pueden determinarse con exactitud, sino que sólo pueden estimarse a partir de una muestra. Por lo tanto, una cuestión esencial es ¿cuáles son o deben ser las características de dicha muestra tales que me permitan hacer inferencias de la muestra a la población?

Los *parámetros* se estiman a través de *estadísticos* calculables a partir de la muestra.

Por ejemplo, Landi et al (1990) decidieron trabajar con una muestra de 600 casos. Con base en esa muestra ellos efectivamente han podido calcular el siguiente par de estadísticos para estimar así los parámetros mencionados:

- 1- la edad media de las personas que leen diariamente algún periódico en la muestra
- 2- el porcentaje de los habitantes de la muestra que tienen TV

Los estadísticos son los que los investigadores conocen, en tanto que los parámetros son los que quieren conocer. Un estadístico es una característica de una muestra y un parámetro es una característica de la población.

Tabla 11
Notación de la población y de la muestra

Medida	Notación de la muestra	Notación de la población
Media	\bar{x}	μ
Proporción	p	π
Desviación estándar	s	σ^2
Varianza	s^2	σ_x^2

Nota: Como pueden observar se emplean letras latinas minúsculas para definir estadísticos de muestra y letras griegas o latinas mayúsculas para representar parámetros de población.

La cantidad de casos que integran una muestra es considerablemente inferior a la de la población, debido a esto, fijado un n (cantidad de casos de la muestra) se puede obtener una cantidad considerablemente grande de muestras posibles (combinación de n valores seleccionados de los N de la población). Ahora bien, es razonable pensar que al tener diferentes muestras de la misma población y calcular los diferentes estadísticos de cada una, esos valores van a diferir de muestra a muestra. Por lo tanto, un estadístico no es un valor fijo, sino que presenta las siguientes características:

- Puede tener varios resultados posibles.
- No se puede predecir de antemano su valor.

Estas son las condiciones que definen a una variable aleatoria.

¿Cuándo hacer estimaciones con base en una muestra?

Gracias a la teoría de las probabilidades, se sabe que el muestreo permite deducir, a partir de los resultados muestrales, las características de la población a la que se aplica. Para que eso sea posible, es necesario que en el proceso de diseño y selección

de la muestra se cumplan determinadas condiciones. De lo contrario, el investigador se encuentra con los límites de los datos: o bien, los datos no son reproducción a escala reducida de la población o su precisión es tan escasa que la inferencia se hace imposible.

Entonces, estimar los parámetros a través de una muestra sólo es posible cuando ésta representa a la población. En la práctica de la investigación, para saber cuándo una muestra es representativa de la población debemos analizar cómo fue extraída; esto es, *analizar el diseño de la muestra y su desarrollo*. Solo algunos procedimientos de selección de muestras proporcionan muestras representativas.

Los procedimientos que se emplean en la selección de los casos constituyen un factor clave si se pretenden hacer inferencias estadísticas sobre características determinadas de la población.

Los procedimientos que permiten hacer inferencias, es decir, estimar parámetros a través de estadísticos, son aquellos que se basan en el uso planeado del azar. Es decir, las muestras probabilísticas.

De este modo, para que la inferencia sea posible, las muestras deben reunir tres condiciones críticas:

1. homogeneidad: todas las unidades deben ser extraídas de la misma población,
2. independencia: las observaciones no deben estar mutuamente condicionadas entre sí, y
3. representatividad: la muestra debe ser el mejor reflejo posible del conjunto del cual proviene.

Muestreos probabilística o aleatorios

La probabilidad es el grado de “confianza” o “creencia” que un individuo (o varios) asigna a la ocurrencia de un evento basándose en la evidencia que se dispone. Es un valor entre 0 y 1. Por ejemplo, la probabilidad de que salga cara o cruz cuando uno arroja una moneda al aire es $0,5$ ($1/2$).

Teniendo en cuenta esta definición, una muestra probabilística es la que se apoya en la teoría de las probabilidades (o el arte de cuantificar la incertidumbre). Su empleo, por lo tanto, origina cierto grado de riesgo y obliga a convivir con el “error muestral” (por no estudiar el total de la población).

En las muestras probabilísticas cada elemento del universo tiene una probabilidad conocida de integrar la muestra y dicha probabilidad no es nula para ningún elemento. Los métodos de muestreo no probabilísticos no garantizan la representatividad de la muestra y por lo tanto no permiten realizar estimaciones inferenciales sobre la población.⁷

Entre los métodos de muestreo probabilísticos más utilizados en investigación encontramos:

⁷ Entre los muestreos no probabilísticos se encuentran: *el muestreo bola de nieve*, *el muestreo con intencional* y *el muestreo por cuotas*. En el primero, el método de selección de casos no es controlado y los integrantes de la muestra se encuentran en ella por propia decisión y sin participación del azar. En el intencional se realiza la selección de casos para cubrir ciertos aspectos particulares que “suponemos” a priori que se deben encontrar presentes en la muestra. Por último, en el muestreo por cuotas se seleccionan los elementos según uno o varios atributos replicando la estructura que se conoce o supone de la población.

- Muestreo aleatorio simple
- Muestreo estratificado
- Muestreo sistemático
- Muestreo polietápico o por conglomerados

Muestreo aleatorio simple.

Selecciona muestras mediante métodos que permiten que cada posible muestra tenga igual probabilidad de ser seleccionada y que cada elemento de la población total tenga una oportunidad igual de ser incluido en la muestra.

La forma más fácil de seleccionar una muestra de manera aleatoria simple es asignarle a cada unidad de análisis de la población un número y realizar un sorteo (por ejemplo con un bolillero) para seleccionar la n cantidad de unidades de compondrán la muestra. Otro modo de realizar esta selección es mediante el uso de números aleatorios. Estos números pueden generarse ya sea con una computadora programada para resolver números o mediante una tabla de números aleatorios (tabla de dígitos aleatorios).

Muestreo sistemático.

En el muestreo sistemático, los elementos son seleccionados de la población dentro de un intervalo uniforme que se mide con respecto al tiempo, al orden o al espacio.

El muestreo sistemático difiere del aleatorio simple en que cada elemento tiene igual probabilidad de ser seleccionado, pero cada muestra no tiene una posibilidad igual de ser seleccionada.

Para la selección de elementos se determina un intervalo k , en función de la cantidad de casos de la población y los de la muestra ($k = N / n$). Se selecciona un número al azar entre 1 y k , este número es r .

Paso siguiente se seleccionan las unidades r ; $r + k$; $r + 2k$; $r + 3k$; Hasta llegar a completar la cantidad de casos necesarios para la muestra.

Por ejemplo: $k = N / n = 1000 / 100 = 10$. Entre 1 y 10 surge por azar el 2. Entonces se elige la segunda unidad de cada uno de los grupos (2; 2 + 10; 2 + 20; 2 + 30; = 2; 12; 22; 32; ...)

Este método puede requerir menos tiempo y algunas veces tiene como resultado un costo menor que el método aleatorio simple.

Muestreo estratificado.

El muestreo estratificado implica la división de la población en grupos relativamente homogéneos llamados estratos y la selección de muestras independientes en cada uno de esos estratos. Usando información relevante acerca de una población, el muestreo estratificado reduce las muestras posibles que pueden seleccionarse a aquellas que proporcionan una mejor representación de la población. Por lo tanto el muestreo estratificado es una técnica que usa tal información para aumentar la eficiencia.

Después, se utiliza uno de estas opciones:

- Seleccionamos aleatoriamente de cada estrato un número específico de elementos correspondientes a la fracción de ese estrato en la población como un todo.

- Extraemos un número igual de elementos de cada estrato y damos peso a los resultados de acuerdo con la porción del estrato con respecto a la población total.

Con cualquiera de estos planteamientos, el muestreo estratificado garantiza que cada elemento de la población tenga posibilidad de ser seleccionado.

Este método resulta apropiado cuando la población ya está dividida en grupos de diferentes tamaños y deseamos tomar en cuenta este hecho (por ejemplo: categorías profesionales de la población).

La ventaja de las muestras estratificadas es que, cuando se diseñan adecuadamente, reflejan de manera más precisa las características de la población de la cual fueron elegidas.

Muestreo por conglomerados.

El muestreo simple al azar, desarrollado anteriormente, presenta dos inconvenientes:

- Cuando cada individuo seleccionado tiene asociado un costo de desplazamiento para poder realizar la entrevista, el muestreo simple al azar puede dispersar la muestra físicamente de tal manera de elevar el costo total de forma considerable. Esta situación es muy común en encuestas a o ares que intentan cubrir grandes áreas geográficas.
- Para poder seleccionar una muestra simple al azar, es necesario disponer de un marco completo de las unidades a interrogar con una identificación física y geográfica para cada una de ellas. Algunas veces surgen situaciones en las que el listado de las unidades o marco, no está disponible o está obsoleto.

Para contener estos dos problemas de naturaleza bien distinta se puede emplear la técnica de muestreo por conglomerados y multietápico. El muestreo multietápico se refiere al proceso de seleccionar una muestra en dos o más etapas sucesivas. Esto implica la existencia de diferentes tipos de unidades muestrales asociadas a cada etapa y a una organización jerárquica de las mismas.

En este procedimiento se forman conglomerados (grupos individualmente representativos de la población), de modo que exista una variación considerable dentro de cada grupo, pero que los grupos sean esencialmente similares entre sí.

En el caso de realizarse en una etapa, se divide a la población en conglomerados y luego se elige uno de ellos al azar. Todos los elementos del conglomerado seleccionado son incluidos en la muestra.

Si se realiza en dos etapas, se divide a la población en conglomerados (en este caso: unidades primarias de muestreo) y luego se elige una muestra de ellos al azar. Dentro de los aglomerados seleccionados se elige una muestra probabilística de unidades.

Relaciones entre el diseño de la muestra y la realización de inferencias

Con base en el supuesto de aleatoriedad en la selección de las muestras, es posible sostener que las estimaciones son *insesgadas* y calcular los errores de muestreo que permiten determinar la *precisión* de dichas estimaciones.

¿Qué significa que una estimación es *insesgada*? Significa que, en promedio, los valores del estadístico obtenidos a partir del muestreo son iguales al parámetro.

La *precisión* hace referencia a la concentración de valores en el muestreo, es decir a la poca variabilidad. La obtención de estimadores - estadísticos -- insesgados y precisos es condición básica para realizar inferencia estadística. Supuesta aleatoriedad y la independencia, la precisión también se relaciona con el tamaño de

la muestra. Es decir, los niveles de desagregación que se pretenden alcanzare en las conclusiones deben corresponderse con las características del diseño muestral. Si esta condición no se cumple lo que posiblemente ocurra es que se supere ampliamente lo que se podría llamar un error permitido. Sin embargo, debemos advertir que cuando el proceso de selección esta sesgado, la utilización de una muestra grande no corrige el error básico, sino que simplemente lo repite a mayor escala.

Para que el muestreo sea probabilístico es necesario que se respete la aleatoriedad y la independencia en todo el proceso de diseño y selección de la muestra, aún en el trabajo de campo. Es necesario respetar la aleatoriedad al elegir y aplicar el método de selección de las unidades de muestreo, pero también es necesario respetar la aleatoriedad a la hora de recoger la información. Es aquí donde las dificultades son mayores.

Si no se cumple la condición de aleatoriedad, se dice que el proceso de selección de las muestras está sesgado. Cuando ocurren *sesgos de selección* la muestra pierde su condición de probabilística. Los sesgos de selección tienen lugar cuando no se respetan rigurosamente las instrucciones referidas a la selección de individuos y se recurren a sustituciones o reemplazos.

La elección de una muestra debe ser un proceso lo más cuidadoso y preciso posible, que seleccione individuos para su inclusión en la muestra de un modo imparcial. De tal forma que se obtenga una muestra representativa.

El sesgo de selección surge en la primera etapa del muestreo;

Existe una segunda etapa en el proceso de selección de unidades de análisis. Después de decidir efectivamente qué personas deben incluirse en la muestra, aún hemos de recoger los datos. Esto es más difícil de lo que parece. Si un gran número de los individuos seleccionados no responde al cuestionario puede crearse una seria distorsión llamada *sesgo de no respuesta*. Los individuos que no responden pueden ser muy diferentes de los que responden. Cuando el porcentaje de los que no contestan es muy alto se produce *el sesgo de no respuesta*.

Anexo 1

¿Cómo se construyen las distribuciones de frecuencias?

Hemos introducido en este Teórico el concepto de *distribución de frecuencias*; sin embargo, me gustaría detenerme un poco en *cómo se construye una distribución de frecuencias*. A pesar de que este proceso habitualmente lo hacemos a través de la computadora, asistidos por programas o paquetes estadísticos con el SPSS, es importante analizar el proceso para comprender cómo el investigador llega a una *distribución de frecuencias*.

Para ejemplificar este proceso, voy a utilizar como insumo una submuestra de 30 casos reconstruida con base en la lectura del texto de Landi et al (1990). Estas variables fueron relevadas en la encuesta sobre públicos y consumos culturales. Las respuestas dadas por los integrantes de la submuestra de 30 casos fueron organizadas en la Tabla 8. Noten que como en toda matriz de datos cada reglón de la tabla representa a una unidad de análisis (encuestado), mientras que cada columna representa a una variable.

Tabla 12
Submuestra de 30 encuestados

Edad	Género	¿Cuál es el medio al que los entrevistados le creen más?
15	Varón	Ninguno
16	Mujer	Diarios
17	Varón	Radios
18	Mujer	Radios
19	Varón	Radios
19	Mujer	TV
20	Varón	Ninguno
24	Mujer	Diarios
27	Varón	TV
28	Mujer	Ninguno
28	Varón	Diarios
30	Mujer	TV
30	Varón	Ninguno
36	Varón	Diarios
37	Varón	Radios
40	Mujer	TV
40	Varón	Ninguno
43	Varón	Diarios
48	Mujer	Revista
49	Mujer	Ninguno
50	Varón	Diarios
50	Mujer	Radios
52	Varón	Radios
54	Mujer	TV
57	Mujer	Ns/nc
58	Varón	Ninguno
58	Varón	Diarios
62	Mujer	Diarios
62	Mujer	Ninguno
62	Mujer	Ninguno

¿Cuántos de los 30 entrevistados son varones? ¿Cuántos dijeron que los diarios son el medio más creíble? ¿Cuántos tienen entre 15 y 17 años?

Empecemos con la variable género. Lo primero que tenemos que hacer es “palotear” el número de hombres y luego el de mujeres (La columna de palotes se incluye en Tabla 13 con propósitos ilustrativos). Los palotes son luego utilizados para construir la distribución de frecuencias presentada en la Tabla 13.

Tabla 13
Distribución de frecuencias de la variable género. Submuestra.

Género	Palotes	Frecuencia (f)	%
Varón		15	50
Mujer		15	50
Total (n)		30	100

Mirando esta distribución de frecuencias rápidamente advertimos que 15 de los 30 entrevistados son varones, es decir, el 50% de la submuestra.

Hagamos lo propio con la edad...La tabla 14 pone de manifiesto que 3 de los encuestados tienen entre 15 y 17 años y que representan el 9,9% del total de la submuestra.

Tabla 14
Distribución de frecuencias de la variable género. Submuestra.

Edad	Palotes	Frecuencia (f)	%
15	I	1	3,3
16	I	1	3,3
17	I	1	3,3
18	I	1	3,3
.	.	.	.
.	.	.	.
.	.	.	.
62		3	10,0
Total (n)		310	100