

• **Introducción al Análisis de Datos**

El análisis de datos de la encuesta tiene como objetivo la detección de grupos variables altamente relacionados, para ello se utilizan los siguientes análisis :

- **Análisis Descriptivo** : ayudará a observar el comportamiento de la muestra en estudio, a través de tablas, gráficos.....

Los resultados recogidos en la muestra se resumen en una matriz de datos **N x M** , en la cual **N** es el número de unidades de análisis utilizadas (número de casos) y **M** es el número de características de dichas unidades , unidades de las que tenemos información.

- **Análisis Exploratorio** : la analización exploratoria pretende partir de un conocimiento profundo y creciente de los datos para , trabajando inductivamente , llegar a un modelo ajustado de los datos. Los pasos en este tipo de análisis son los siguientes :
- Análisis de cada una de las variables incluidas en la matriz de datos. Los datos se agrupan de un modo rápido y a ser posible gráfico , las técnicas mas utilizadas son :

- **Tronco y hoja**
- **La caja**

Ambas pretenden :

- Conocer la variable analizada para determinar si su distribución es simétrica o no.
 - Poder descubrir valores extremos y analizarlos antes de poder pasar al análisis multivariante.
- **Transformación de los datos** : la transformación persigue la consecución de una distribución aproximada a la normal. Tipos de transformación :
 - **Lineales** : suma , resta , división , multiplicación , cambia los valores brutos (datos obtenidos) de la variable sin alterar nada mas.
 - **No lineales monotónicas** : cambian los valores originales y tambien sus distancias pero no el orden
 - **No lineales no monotónicas** : similar a la anterior pero no altera el orden..
 - **Análisis Confirmatorio / Explicativo** : la mayor parte de las técnicas tradicionales de análisis estadístico de los datos tienen un carácter deductivo confirmatorio. De todas las técnicas de análisis confirmatorio la más útil para el sociólogo es aquella que parte del análisis de variables , entre las que cabe distinguir : **nominales** y de **intervalo o de razón** . En la investigación sociológica las de tipo ordinal son muy frecuentes , pero las técnicas de análisis escasas , de modo que se tratan como variables de intervalo o razón , labor muy complicada o como nominales.
 - **Investigación con Variables Nominales**

En la investigación con variables nominales se utilizan técnicas como el análisis de varianza o factorial.

Dentro de las técnicas multivariantes de análisis de variables nominales existen dos de caracter especial :

- Los coeficientes χ^2 J. Davis ha perfeccionado un modelo que se basa en las diferencias porcentuales. Los coeficientes χ^2 son diferencias entre proporciones y se utilizan como indicadores del impacto causal de unas variables sobre otras.

- Modelos <log–linear> explican la probabilidad de que una persona elegida al azar presente una determinada combinación de categorías/ niveles del conjunto de variables de clasificación utilizados

6. Investigación con variables de intervalo o de razón : en estas se utiliza el análisis de regresión : puede utilizarse siempre que la variable dependiente sea una variable de intervalo.

2. Análisis de los Datos de Encuesta

El análisis de los datos de la encuesta , como cualquier otro tipo de datos de interés científico , ha de guardar relación con el problema de conocimiento que se trata de esclarecer y con la métrica de la información empírica que se tiene entre manos , es decir lo primero que se debe realizar en una encuesta no es ver que dicen los datos sino que dicen en relación con el problema que se plantea y las hipótesis que uno se había planteado previamente.

Una serie de conclusiones importantes sobre los datos de una encuesta son :

- La cantidad y calidad del conocimiento que se desea obtener sobre un problema no está necesariamente en función del tamaño de la muestra empleada para hacer una encuesta.
- Ni el margen global de error en los resultados de una encuesta disminuye necesariamente aumentando el tamaño de la muestra.
- Los datos de la encuesta son mas útiles cuanto mayor sea la posibilidad de poderlos comparar con otros similares y anteriores en el tiempo o procedentes simultáneamente de otras poblaciones.

Tipos de Análisis :

Análisis de Correlación : Con los datos mas simples de respuestas a varias preguntas se pueden fabricar medidas combinatorias de los mismos en forma de índices o escalas.

Estas medidas constituyen la **operacionalización** de los conceptos, es la referencia empírica reducida a valores empíricamente manejables.

Los **conceptos** son elaboraciones teóricas con las que definimos un determinado aspecto de la realidad y de la que nos servimos para expresar teorías e hipótesis.

Las **variables** : son dimensiones de la realidad que deseamos estudiar.

Los **indicadores** : son expresiones numerables de las distintas dimensiones de un concepto.

Con los datos indexados podemos hacer análisis de correlación de diferentes tipo.

La correlación entre dos o mas variables expresa la extensión en que una incide en la otra intersección, la intersección de ambas, su grado de asociación... . Pero no indica si a partir de los valores de una de ellas podemos predecir los valores de la otra , esto se consigue mediante el **coeficiente de regresión** que indica si la correlación entre variables es tal que a partir de una podemos calcular los valores de otra. Todo esto lo podemos realizar con medidas de la estadística no paramétrica no pueden ser calculadas.

A partir del Análisis de Correlación se puede trabajar con distintos modelos de análisis multivariable :

Un conjunto de modelos de análisis podría denominarse como el del **análisis causal** : aquí se incluirían los **análisis de cadena** (< **path análisis**>) , el de **regresión múltiple** y el de **segmentación por arborescencia** (<**tree análisis**>) el cual es una técnica multivariable que produce una serie de agrupamientos dicotomizados en un modelo que se asemeja a las ramas de un arbol. La división de las cosas en grupos dicotomizados la

realiza el **análisis de segmentación** de modo que en ellos quede explicada de la mejor forma posible la variación de una cierta variable que llamamos dependiente , respecto a otras que llamamos independientes. La división en grupos dicotomizados e debarlizar en cada paso de tal manera que la variación en ellos sea mínima, que sean lo mas homogéneos posibles al mismo tiempo que la variación entre si sea la mayor posible, que los dos grupos resultantes sean lo mas heterogéneos posibles entre si.

Otro conjunto estaría formado por distintos modelos que implican conglomerados de variables o asociaciones entre los que se encuentran :

- **El análisis factorial** : es una forma de análisis de matrices. El objeto de este método es resumir la información obtenida de las variables iniciales expresando las mismas como combinación lineal de otras variables no observables (subyacentes) denominadas factores. Se busca una estructura de interrelación en muestras variables. Para decidir si esta análisis factorial es factible utilizaremos ciertas medidas como pude ser la :
 - ◊ **Kaiser–Meyer–Olkin** : basada en los coeficientes de correlación parcial situando su rango de variación entre 0 y 1.
 - ◊ **Medida de adecuación muestral** : la cual nos indica hasta que punto debe entrar una variable en el análisis factorial.
- **El análisis Tipológico** : es un análisis de estructuras latentes , nos permite identificar dentro de una población determinada un cierto número de grupos, los cuales tienen una igualdad características en una multiplicidad de dimensiones.
- **El del espacio mas pequeño.**

2. Análisis de los Datos de Encuesta

El análisis de los datos de la encuesta , como cualquier otro tipo de datos de interés científico , ha de guardar relación con el problema de conocimiento que se trata de esclarecer y con la métrica de la información empírica que se tiene entre manos , es decir lo primero que se debe realizar en una encuesta no es ver que dicen los datos sino que dicen en relación con el problema que se plantea y las hipótesis que uno se había planteado previamente.

Una serie de conclusiones importantes sobre los datos de una encuesta son :

- La cantidad y calidad del conocimiento que se desea obtener sobre un problema no está necesariamente en función del tamaño de la muestra empleada para hacer una encuesta.
- Ni el margen global de error en los resultados de una encuesta disminuye necesariamente aumentando el tamaño de la muestra.
- Los datos de la encuesta son mas útiles cuanto mayor sea la posibilidad de poderlos comparar con otros similares y anteriores en el tiempo o procedentes simultáneamente de otras poblaciones.

Tipos de Análisis :

Análisis de Correlación : Con los datos mas simples de respuestas a varias preguntas se pueden fabricar medidas combinatorias de los mismos en forma de índices o escalas.

Estas medidas constituyen la **operacionalización** de los conceptos, es la referencia empírica reducida a valores empíricamente manejables.

Los **conceptos** son elaboraciones teóricas con las que definimos un determinado aspecto de la realidad y de la que nos servimos para expresar teorías e hipótesis.

Los **variables** : son dimensiones de la realidad que deseamos estudiar.

Los **indicadores** : son expresiones numerables de las distintas dimensiones de un concepto.

Con los datos indexados podemos hacer análisis de correlación de diferentes tipo.

La correlación entre dos o mas variables expresa la extensión en que una incide en la otra intersección, la intersección de ambas, su grado de asociación... . Pero no indica si a partir de los valores de una de ellas podemos predecir los valores de la otra , esto se consigue mediante el **coeficiente de regresión** que indica si la correlación entre variables es tal que a partir de una podemos calcular los valores de otra. Todo esto lo podemos realizar con medidas de la estadística no paramétrica no pueden ser calculadas.

A partir del Análisis de Correlación se puede trabajar con distintos modelos de análisis multivariable :

Un conjunto de modelos de análisis podría denominarse como el del **análisis causal** : aquí se incluirían los **análisis de cadena** (< **path análisis**>) , el de **regresión múltiple** y el de **segmentación por arborescencia** (<**tree análisis**>) el cual es una técnica multivariable que produce una serie de agrupamientos dicotomizados en un modelo que se asemeja a las ramas de un arbol. La división de las cosas en grupos dicotomizados la realiza el **análisis de segmentación** de modo que en ellos quede explicada de la mejor forma posible la variación de una cierta variable que llamamos dependiente , respecto a otras que llamamos independientes. La división en grupos dicotomizados e deberalizarse en cada paso de tal manera que la variación en ellos sea mínima, que sean lo mas homogéneos posibles al mismo tiempo que la variación entre si sea la mayor posible, que los dos grupos resultantes sean lo mas heterogéneos posibles entre si.

Otro conjunto estaría formado por distintos modelos que implican conglomerados de variables o asociaciones entre los que se encuentran :

- **El análisis factorial** : es una forma de análisis de matrices. El objeto de este método es resumir la información obtenida de las variables iniciales expresando las mismas como combinación lineal de otras variables no observables (subyacentes) denominadas factores. Se busca una estructura de interrelación en muestras variables. Para decidir si esta análisis factorial es factible utilizaremos ciertas medidas como puede ser la :

◇ **Kaiser–Meyer–Olkin** : basada en los coeficientes de correlación parcial situando su rango de variación entre 0 y 1.

◇ **Medida de adecuación muestral** : la cual nos indica hasta que punto debe entrar una variable en el análisis factorial.

- **El análisis Tipológico** : es un análisis de estructuras latentes , nos permite identificar dentro de una población determinada un cierto número de grupos, los cuales tienen una igualdad características en una multiplicidad de dimensiones.
- **El del espacio mas pequeño.**

El Análisis de Contenido

Definición : El análisis de contenido es una técnica de investigación para la descripción objetiva, sistemática y cualitativa del contenido manifiesto de la comunicación (Berelson 1952) , esta es la definición mas

importante pero también existen otras como puede ser la de Krippendorff (1980) según la cual el análisis de contenido es una técnica de investigación para hacer inferencias reproducibles y válidas de los datos.

Aplicaciones del análisis de Contenido

Los objetivos del análisis de contenido son tres :

- La descripción precisa y sistemática , de las características de una comunicación.
- La formulación de inferencias sobre contenidos exteriores al contenido de la comunicación.
- La prueba de hipótesis para su verificación o rechazo.

Para Janis (1965) todo análisis de contenido debería tener como fin el someter a prueba alguna relación entre características del contenido y otras variables.

Dada esta definición se pueden obtener diferentes tipos de análisis de contenido :

- **Análisis de contenido semántico** : clasificación de los signos según su significado : ejemplo número de referencias al Fc.Barcelona.
- **Análisis de designaciones** : número de veces que aparecen referencias a determinados objetos.
- **Análisis de atribución** : proporciona la frecuencia con que aparecen ciertas caracterizaciones : número de veces que aparece una característica.
- **Análisis de Afirmaciones** : proporciona la frecuencia en la que ciertos objetos aparecen caracterizados de una determinada manera.
- **Análisis de contenido pragmático** : clasificación de signos según sus causas o efectos probables.
- **Análisis de vehículo del signo** : clasificación del contenido según las proporciones psicofísicas de los signos.

El contenido tradicional del análisis de contenido es la descripción; se centra en el contenido y no en hipótesis , por lo tanto el objetivo principal es entonces la descripción de ese contenido, de su fondo y de su forma. (suelen ser investigaciones de carácter exploratorio o de índole histórica).

La clasificación más completa del análisis de contenido de carácter descriptivo es la realizada por Berelson (1952) el cual utilizó el criterio de la < función > del análisis de Holsti (1969) es la siguiente :

- **Análisis de las características del contenido en cuanto a su sustancia o fondo :**
- Para descubrir tendencias y cambios en el contenido de la comunicación.
- Para averiguar la evolución de los intereses y de las actividades del pensamiento erudito.
- Para revelar diferencias internacionales en el contenido de la comunicación.
- Para realizar comparaciones entre diferentes medios de comunicación y entre diferentes niveles de comunicación.
- Para analizar el contenido de la comunicación en términos de sus objetivos explícitos o implícitos.
- Para elaborar y poner en vigor ciertos estándares de comunicación así como examinar el contenido de las comunicaciones en términos de tales estándares.
- Para ayudar en operaciones técnicas de investigación, como en la codificación de datos obtenidos en entrevistas u observaciones no es-estructuradas.
- Para relaciones características conocidas de las fuentes de la comunicación con los mensajes que tales fuentes producen.
- Para relacionar características conocidas de la audiencia con mensajes creados para ella.
- **Análisis de las características del contenido en cuanto a su forma :**
- Para revelar técnicas de propaganda o persuasión.
- Para medir la legibilidad (interés, amenidad.....) de material impreso.
- Para descubrir rasgos estilísticos en lenguajes, en periodos históricos en tipos de discurso y en

autores individuales.

El análisis de contenido de carácter descriptivo se caracteriza por producir determinados tipos de generalizaciones, las cuales muchas provienen de la comparación de contenidos de comunicaciones de diferente origen (diarios italianos, franceses...), o comunicaciones de único origen pero diferente tiempo, o comunicaciones de único origen pero de situaciones distintas, en otras ocasiones son generalizaciones que proceden de la comparación del contenido de una o más comunicaciones y finalmente las que proceden del análisis de relación o asociación entre dos o más variables, destacando aquí el análisis de contingencia: el de la aparición simultánea de determinadas variables.

Otro objetivo del análisis de contenido es la realización de **inferencias**, cuando el contenido de los mensajes es analizado para apoyar conclusiones no relacionadas con el contenido, entonces adquiere la investigación un carácter explicativo o inferencial.

Hay dos categorías de **inferencias**:

- Las que se refieren al origen, causas, condiciones, antecedentes de la comunicación y especialmente al autor de la misma, estas inferencias responden a las preguntas < quién > y < por qué >.
- La segunda está formada por las inferencias que se refieren a efectos o resultados de la comunicación y el análisis responde al interrogante < con qué efecto >.

Combinando las aportaciones de Berelson y Holsti se obtiene la siguiente clasificación del análisis de contenido de carácter inferencial:

- **Análisis para realizar inferencias acerca de las causas, de los productores o de los antecedentes de la comunicación:**
- **Para identificar las intenciones y otras características de los creadores de las comunicaciones.**
- **Para analizar rasgos psicológicos de individuos o determinar el estado psicológico de personas o grupos.**
- **Para detectar la existencia de propaganda.**
- **Para obtener información política y militar.**
- **Para determinar actitudes, intereses, valores de grupos o poblaciones, así como el cambio en los mismos que haya podido tener lugar.**
- **Para proporcionar pruebas jurídicas.**
- **Para decidir disputas sobre paternidad literaria.**
- **Análisis para realizar inferencias de las consecuencias o efectos de la comunicación:**
- **Para determinar qué actitudes y comportamientos son consecuencia de la comunicación.**
- **Para analizar la corriente de información.**
- **Para medir la legibilidad de textos.**

De todo esto dicho anteriormente se extraen dos conclusiones:

- el análisis de contenidos puede ser defectuoso por aplicación errónea de las técnicas de análisis, en cuyo caso la descripción del contenido de las comunicaciones estará más o menos llena de errores.
- el análisis de contenido puede ser defectuoso aun cuando la descripción del contenido sea correcta, por aplicación errónea de las técnicas de inferencia o interpretación.

De la clasificación de Krippendorff se extraen dos conclusiones:

- que el objetivo del análisis de contenido parte de dos ideas interrelacionadas.
- que el análisis de contenido puede adoptar diversas formas, entre las que destacan:
- **Sistemas:** aparatos conceptuales que se utilizan para describir una porción de realidad.

- **Estándares** : criterios con los que se comparan objetos , para determinar cuan buenos son o de qué tipo son.
- **Índices** : variables cuya significación en una investigación depende del grado en que verse correlacionado con otros fenómenos.

Elementos componentes del Análisis de contenidos

Los Datos : Es una unidad de información grabada en un medio resistente y duradero (papel, cinta...) que se distingue de otros datos en que es analizable por medio de técnicas explícitas siendo significativos para un problema determinado.

El Muestreo

Las unidades de Análisis : son los elementos de la comunicación en que se va a centrar el análisis.

Las unidades de muestreo : son las diversas partes de la realidad sometida a observación que el investigador considera como independientes entre si pudiendo ser objeto de análisis.

Tipos :

- **Artificiales** : bloques de líneas, columnas....
- **Naturales**: editoriales, artículos, carteles....

La unidad de registro : es la mínima porción del contenido que el investigador aísla y separa por aparecer allí uno de los elementos, símbolos, slogans , que el investigador considera significativos.

La unidad contexto : es la porción de la unidad de muestreo que tiene que ser examinada para poder caracterizar una unidad de registro.

El Registro : proceso por el cual cada unidad de registro es codificada, descrita en forma analizable.

La codificación es el proceso por el cual los datos brutos son transformados y clasificados en categorías.

Categorías :

- **Materia , tema , asunto** : da respuesta al interrogante : ¿ De que se trata la comunicación?
- **Dirección o punto de vista** : de este derivan categorías de tratamiento que puede ser desfavorable o favorable.
- **Criterio o valor** : determina la base's que sirven para hacer la clasificación en términos de dirección o punto de vista, la base sobre la que descansa la presentación o el tratamiento del asunto o materia por el autor de la comunicación.
- **Autoridad Frente a Prueba** : de aquí surgen categorías que identifican a las personas, grupos, instituciones....
- **Objetivo Meta** : las categorías que se derivan de este se refieren los objetivos que se desean alcanzar o que se consideran deseables. El interrogante al que da respuesta es : ¿ Qué es lo que la gente desea, que persigue, cuales son sus ambiciones?.
- **Medio o método** : como se consiguen los objetivos que se consideran deseables.
- **Actores** : son los grupos , personas , actores, que ocupan posiciones centrales o que llevan a cabo determinadas acciones , es decir , clasifica aquellos que en el texto sometido a análisis son representados como realizadores de determinados actos.
- **Características personales y rasgos de carácter** : en este principio de clasificación se incluyen categorías que describen características personales (edad, sexo, raza...)
- **Origen** : clasifica el lugar de origen de las comunicaciones.

- **Destino** : clasifica el grupo al que la comunicación está especialmente dirigida.
- **Formas de la afirmación** : este principio elabora subcategorías como la distinción entre proa y poesía , entre novela y ensayo.....
- **Intensidad** : clasifica el componente emocional de la comunicación, el entusiasmo, excitación.....
- **Mecanismo o recurso** : se utiliza con fines propagandísticos o persuasorios o tendenciosos.

La Validez en el Análisis de Contenido : la validez de un análisis se juzga por el grado en que los resultados que produce son compatibles con otras pruebas u testimonios importantes y pertinentes.

La práctica muestra dos tipos de validez :

- **Interna** : la que descansa en la existencia de una relación lógica entre indicadores , categorías y resultados de análisis.
- **Externa** : la que se apoya en una relación empírica. Los tipos de validez externa que mas se adecuan al análisis de contenido son :
 - ◊ **La validez Pragmática** : cuando el interés se centra en la utilidad de los resultados del análisis para indicar o precisar algún comportamiento o alguna característica específica de individuos, grupos..... Esta validez puede ser **concurrente** : cuando el investigador está interesado en alguna característica del presente o **predictiva** : cuando está interesado en alguna característica del pasado.
 - ◊ **Validación de rasgos abstractos** : se utiliza para hacer referencias respecto a algún rasgo abstracto y multidimensional que se supone está reflejado en los resultados del análisis de contenido.
 - ◊ **Validez del Contenido** : se dice que la muestra o el contenido es válido si su composición corresponde a la composición de universo de posibles indicadores de la característica, atributo o comportamiento de que se trata.
 - ◊ **Validez Semántica** : podrá comprobarse la validez semántica comprobando si son similares o no los resultados de dos reparticiones diferentes del mismo conjunto de unidades de análisis, una realizada analíticamente y otra según el criterio externo.