

Semana 1

Presentación *Valentín Vergara Hidd*

En este documento mostraré algunos conceptos relevantes para comenzar a explorar SPSS, o cualquier software estadístico.

Antes de comenzar

Es fundamental que manejemos los mismos conceptos. Existen muchas formas de abordar la estadística, especialmente entre diferentes disciplinas que la utilizan como una **herramienta** para responder preguntas que consideran fundamentales.

Considerando lo anterior, hay ciertos conceptos vinculados a la metodología de investigación o a la epistemología ([Samaja, 2004](#)) que es pertinente revisar antes de continuar con los aspectos más técnicos del software. En primer lugar, les presentaré las *matrices de datos*, que son la forma más genérica en la que podemos mostrar datos *analizables* estadísticamente. También, son la forma más común en que dichos datos se encuentran a disposición de investigadores.

El segundo concepto relevante tiene que ver con la forma en que creamos modelos, no sólo estadísticos. Los modelos, en general, responden a una aproximación incompleta e imperfecta de la realidad, cuya principal función es **simplificar** objetos que de otra forma resultarían muy complejos para ser analizados. Entre ellos, la estadística es una entre muchas herramientas que permiten modelar. Por tanto, existen algunas limitaciones que es muy importante conocer de antemano.

Matrices de datos

Entenderemos matriz como un grupo de elementos ordenados en filas y columnas. Por ejemplo, la Matriz **A** se puede representar como:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

En la matriz, cada elemento a_{ij} tiene dos subíndices. El primero i , indica la posición del elemento respecto a las filas; en tanto que el segundo, j , indica la posición del elemento respecto a las columnas. Este orden de elementos facilita el proceso de *traducir* la experiencia espontánea a una descripción *protocientífica*, que a su vez se convertirá en el material básico de la experiencia científica.

En este punto, es importante aclarar que la traducción entre experiencia y descripción se puede dar de otras formas, pero a lo largo de cientos de años de ejercicio científico, hemos *consensuado* una forma de entender el análisis de datos que descansa en esta arbitraria traducción.

Si pensamos en los datos que observamos sobre cualquier fenómeno, y que intentaremos **ordenar** para entender dicho fenómenos a cabalidad, usualmente operamos de la siguiente forma (Samaja, 2004):

Los casos individuales que nos entregan los datos (individuos, países, muestras, etc.) se agregan como **filas** en la matriz (subíndice i); y corresponde, en términos epistemológicos, a las **unidades de análisis**.

Los atributos que observamos en cada caso, o bien, las **variables** que medimos/observamos; se encuentran en las **columnas** de la matriz (subíndice j).

Los datos que son valores de una **variable** en un **caso** determinado, son los elementos a_{ij} de la matriz. A partir del conjunto de los datos, vamos a intentar extraer información utilizando el menor número posible de transformaciones.

En general, quienes trabajamos de forma cotidiana con datos estamos acostumbrados a verlos ordenados de esta forma. Por ejemplo, una planilla de cálculo, una base de datos, incluso los formularios que llenamos en Internet, terminan convertidos en una estructura de datos que es una matriz. En el caso particular de SPSS, esto es explícito, ya que la primera ventana que se abre al iniciar cualquier conjunto de datos es una matriz con todos los elementos descritos.

El siguiente paso lógico es ver qué hacemos con estos conjuntos de datos, una vez que tenemos una idea más o menos compartida sobre lo que contienen.

Modelos

Sabemos que la estadística se utiliza en prácticamente todos los campos científicos. Los físicos han creado modelos estadísticos muy sofisticados para entender fenómenos a macro o micro escala, difícilmente observables. Los economistas han refinado modelos que permiten entender mercados, precios, o incluso detectar burbujas (aunque no los apliquen demasiado). En ambos ejemplos, la estadística es una herramienta que permite crear un **modelo** de la realidad. Es una representación imperfecta y mucha más simple, que permite entender como funciona un proceso que es órdenes de magnitud más complejo, especialmente si se observa directamente (en caso que se pueda).

Estos modelos, particularmente son **modelos estadísticos**, que intentan representar la contingencia de la realidad en un modelo abstracto que nos permita observar alguno (o algunos aspectos) de dicha realidad que es muy compleja para analizar en su totalidad.

Una forma interesante de ejemplificar lo anterior es a través de una analogía presentada por Field (2018): imaginen a un ingeniero que desea construir un puente para atravesar un río. La construcción no empieza de inmediato, sino que hay todo un procedimiento previo en el que se reúnen datos necesarios en términos de los materiales, las propiedades estructurales del puente y otras experiencias de puentes para atravesar otros ríos. Una vez reunidos estos datos, probablemente lo que el ingeniero haría sería construir un modelo del puente: un modelo a computador o una maqueta a escala. Construir el puente antes de hacer un modelo sería peligroso y posiblemente extremadamente caro. Este modelo difiere en varios aspectos de la versión final del puente: la más importante es su tamaño. Aún así, la idea es que este modelo sea una *representación* de lo que será el puente de verdad una vez construido.

Lo que se hace en estadística es algo similar. Se construyen modelos que permiten acceder a una realidad compleja a través de simplificaciones abstractas que tiene una lógica particular y que

tiene coherencia interna. Tal como los ingenieros, aquellos que usan modelos estadísticos esperan que sus modelos sean lo más parecido posible a la situación que están modelando. A veces hay que escoger entre distintas alternativas a aquella que realmente se acerque a una representación de la realidad. Retomando el ejemplo del puente, sería algo así:

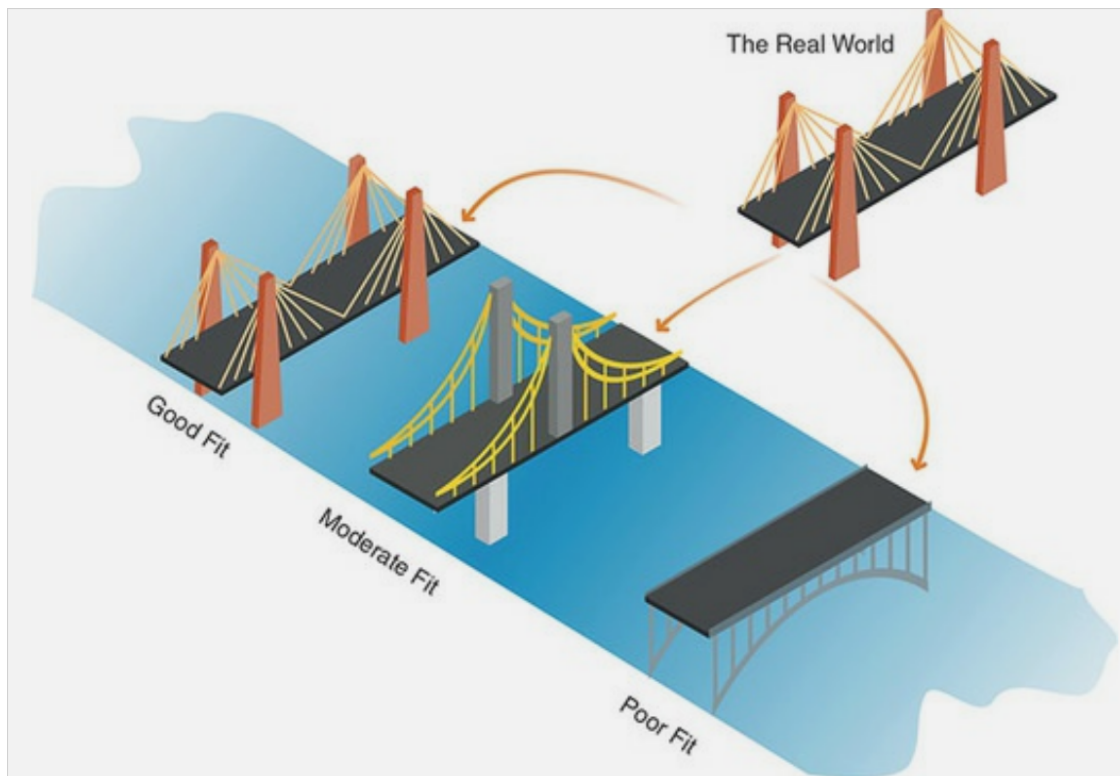


Figura 1: La analogía entre modelos estadísticos y puentes. (Field, 2018), página 96

Obviamente, no todos los modelos son iguales. El problema es mucho más complejo, porque a pesar que evidentemente el modelo de la izquierda (*good fit*) es una mejor representación de la realidad, es el que toma más tiempo y más recursos construir. Si sólo se quiere conocer un aspecto determinado de la realidad, puede que el modelo de la derecha (*poor fit*) sea el menos informativo en términos globales, pero el más útil para un propósito específico. El problema, entonces, requiere optimizar el nivel de detalle del modelo, sujeto a la profundidad del análisis y especialmente a la pregunta de investigación. Hacia el final de este curso vamos a construir modelos estadísticos y volveremos a pensar en estos temas.

Para finalizar esta presentación, quisiera comentar un elemento que muchas veces se pasa por alto: no sólo se pueden crear modelos con estadística. Podemos crear modelos computacionales (simulaciones), modelos puramente matemáticos, podemos representar la realidad por medios narrativos, podemos usar una caricatura, etc. Lo que todos tienen en común es que son **simplificaciones** de la realidad, de modo que podamos entender algún aspecto muy acotado de ella. No hay que perder de vista este punto, los modelos jamás cuentan toda la historia, sólo una parte muy acotada de ella.

Referencias

Field, Andy. 2018. *Discovering statistics using IBM SPSS statistics*. London Thousand Oaks, California: SAGE Publications.

Samaja, Juan. 2004. *Epistemología y Metodología: elementos para una teoría de la investigación científica*. Editorial Universitaria de Buenos Aires, 4 edition.