

---

### RELACIÓN ENTRE LAS VARIABLES

---

El origen de las variables se encuentra en el nivel exploratorio, veamos un ejemplo: hace cientos de años atrás, se describió una enfermedad a la cual hoy conocemos con el nombre de la diabetes. En esa época, se detectó que había un conjunto de síntomas que eran compartidos por algunas personas y que causaban la muerte. En ese momento se mencionó a este conjunto de síntomas como: polidipsia o mucha sed, poliuria o micción frecuente y polifagia como un incremento del apetito.

Este es el origen de las variables, porque partimos de un fenómeno, una condición que afectaba a la personas y que los conducía a la muerte; así que había que describirlo, pero antes tenía que ser identificado. Este es el rol de la investigación exploratoria: identificar un conjunto de características en una población; estas características se convertirán, más adelante, en variables y tendrán que ser analizadas desde el punto de vista estadístico.

Polidipsia, poliuria y polifagia, se resumen en un concepto denominado diabetes. **Aquí nace la variable de estudio** que viene a ser la diabetes, variable que se mantendrá constante a lo largo de toda la línea de investigación, ira adoptando nombres distintos en los diferentes niveles investigativos, pero siempre será la misma variable, es la que representa el problema en estudio.

## 1. Nivel investigativo descriptivo

### 1.1. Las variables descriptivas

Clásicamente el nivel descriptivo se caracteriza por ser univariado, con el término de univariado nos estrictamente al análisis estadístico que vamos a desarrollar, porque además de la variable de estudio, siempre tendremos que caracterizar a nuestra población. Imagina que estás realizando un estudio de prevalencia de diabetes en tu ciudad. La variable de estudio será la diabetes pero, además de describir cuán frecuente es la diabetes en esta población, tendrás que caracterizar a la población.

Es preciso caracterizar a la población en términos de edad y sexo, que son las variables epidemiológicas que no deben faltar en ningún estudio y menos en los estudios descriptivos; pero así como has descrito las características de edad y sexo, puedes describir también otras condiciones como la ocupación, la actividad física, los hábitos nocivos, y todas las características que, creas están relacionadas a tu variable de estudio.

Como es lógico, mientras más experiencia tengas dentro de tu línea de investigación, más variables descriptivas tendrás y, en esta ocasión, sí vale la pena indicar que son varias, porque son muchas las características que se pueden describir alrededor de la variable de estudio, por eso, a estas características se les denomina **variables de caracterización**.

No hay una fórmula matemática para calcular el número de variables de caracterización, se trata del criterio del investigador, parten de su experiencia, son planteamientos empíricos. Mientras mayor sea el número de variables descriptivas que podamos identificar en este nivel, más éxito tendremos en el transcurso de nuestra línea de investigación.

## 1.2. La variable de interés

En el nivel investigativo descriptivo, la variable más importante del trabajo de investigación, lleva el nombre de variable de interés, se trata de la variable de estudio con un nombre distinto, esto es por la función o rol que debe cumplir esta variable en el nivel descriptivo y recibirá otros nombres en los otros niveles de la investigación.

La variable de interés o variable de estudio en un estudio descriptivo nos permite, por ejemplo, clasificar la investigación. Recuerda que existen los estudios retrospectivos y los estudios prospectivos. Esta clasificación de los estudios se realiza a partir de la **planificación de las mediciones** que se ejecuta sobre la **variable de estudio**.

Las variables descriptivas o variables de caracterización, no se utilizan para clasificar a los estudios, entonces, cuando de clasificar a los estudios se trata nuestra atención debe estar enfocada solamente en una variable, en la variable de estudio; pero que en el nivel descriptivo por su función se denomina variable de interés.

Por otro lado, según el **número de mediciones** que se realiza sobre la variable de interés o variable de estudio; si realizamos una sola medición el estudio es transversal; pero si ejecutamos más de una medición, el estudio será longitudinal. Las mediciones únicas o múltiples pueden ser realizadas por el investigador o por un tercero y luego el investigador copia la información, a partir de los registros. De modo que la clasificación de los estudios en retrospectivo y prospectivo es totalmente independiente a la clasificación de los estudios en transversal y longitudinal.

## 2. Nivel investigativo relacional

### 2.1. Las variables asociadas

Con el término asociar queremos resumir todos los tipos de relación que podamos plantear entre dos variables. Nos encontramos en el nivel investigativo relacional caracterizado por plantear la relación entre dos variables. El **estudio relacional** no es un estudio de causa y efecto, por cuanto la asociación solo es el punto de partida de la causalidad, entonces no podemos hablar de variable independiente ni de variable dependiente.

En el estudio relacional, podemos encontrar dos variables analíticas y desde el punto de vista analítico no hay diferencia entre una y otra, porque no estamos planteado una relación causal. Si planteamos un estudio de asociación entre la obesidad y la hipertensión arterial, encontraremos que se trata de dos eventos asociados, pero que ninguno de ellos representa la causa del otro, simplemente son variables asociadas.

Por otro lado, el término correlación se aplica a la relación entre dos variables numéricas e implica que las variaciones de las unidades en un individuo **se correlacionan** con las unidades de otra característica en el mismo individuo. Por ejemplo: cuando las personas tienen más alta la presión arterial en mmHg, es posible que también tengan más altos los niveles de colesterol medidos en mg/dl.

Como es lógico ninguna de estas dos características podrá recibir el nombre de variable independiente ni de variable dependiente, simplemente se llamarán variables asociadas, es que no hay una relación de causalidad ni se cumplen los criterios de causalidad que debe existir en una relación causa-efecto. Es claro que el rol analítico que tienen las observaciones o evaluaciones en este nivel investigativo es el mismo.

## 2.2. La variable de supervisión

El nivel investigativo relacional, se caracteriza por plantear la relación empírica entre dos variables. La relación se plantea bajo una hipótesis con fundamento empírico, que nace de la experiencia subjetiva del investigador, y por eso algunos investigadores dicen que no tiene fundamento. Las dos variables analíticas que participan en el nivel relacional tienen la misma jerarquía, por lo tanto, el rol que van a cumplir en el análisis estadístico es exactamente el mismo.

Desde el punto de vista investigativo, una de las dos variables participantes, es la **variable de estudio** y se le denomina variable de supervisión. No se trata de la variable dependiente, porque no estamos realizando un estudio de causa y efecto. La variable de estudio, es la que caracteriza a nuestra línea de investigación.

Si planteamos un estudio de relación entre el rendimiento académico y la depresión, entonces, la variable de supervisión identifica al campo del conocimiento en donde se desarrolla el estudio. En un estudio de relación entre el rendimiento académico y la depresión; si el estudio está siendo llevado a cabo por un psiquiatra, entonces, su variable de supervisión será la depresión; pero si este mismo estudio es realizado por un licenciado en Educación, su variable de supervisión será el rendimiento académico.

Desde el punto de vista metodológico enfocamos nuestro esfuerzo en la medición de la variable de supervisión y desde el punto de vista analítico, relacionamos a todas las demás variables, con la variable de supervisión evitando así trabajar con matrices de correlación.

### **3. Nivel investigativo explicativo**

#### **3.1. Las variables independientes**

El nivel explicativo se caracteriza por pretender demostrar relaciones de causalidad entre dos variables; entonces aparecen las variables independientes a partir de las variables asociadas del nivel investigativo anterior, pero para transformar una variable asociada en variable independiente hay que seguir una serie de pasos y recomendaciones.

Supongamos que en un estudio relacional existe relación entre la variable asociada y la variable de supervisión, entonces, existen tres posibilidades:

- La variable asociada es el factor de riesgo de la variable de supervisión.
- La variable asociada es la consecuencia de la variable de supervisión.
- La variable asociada y la variable de supervisión son consecuencia de una tercera variable.

Solamente en el caso de que la variable asociada sea un factor de riesgo de la variable de supervisión, se convierte en variable independiente, uno de los requisitos que debemos identificar en las variables independientes es que debieron aparecer antes de la variable dependiente, esto es muy fácil de demostrar cuando hacemos experimentos, pero no todos los estudios explicativos son experimentales.

Desde el punto de vista analítico, uno de los recursos que utilizamos para asegurarnos de que la variable independiente realmente influya sobre la variable dependiente es el análisis estadístico multivariado. Con el cual podremos descartar las asociaciones aleatorias, espurias o casuales que se hayan detectado en el nivel investigativo anterior.

### **3.2. La variable dependiente**

En un estudio de causa y efecto, la variable dependiente corresponde, al efecto o al resultado, es la variable sobre la cual tratamos de demostrar causalidad, se trata de la variable de estudio, es la que mide o describe el problema que generó la línea de investigación; el objetivo del estudio es llegar a conocer las situaciones que hacen que se produzca el problema.

Muchos investigadores utilizan de manera incorrecta los términos de variable independiente y variable dependiente. Esta terminología es exclusiva del nivel explicativo. Los que desconocen estos conceptos suelen llamar variable dependiente a la variable de supervisión del nivel relacional; y también a la variable endógena del nivel predictivo. La confusión se origina porque en todos los casos se trata de la variable de estudio.

El primer principio para demostrar causalidad, es encontrar una relación entre la variable independiente y la variable dependiente, que no sea aleatoria, que no se deba a la casualidad, sino que realmente una esté produciendo la otra. Una forma certera de hacerlo es manipulando la causa para producir el efecto. Se trata de un experimento, uno de los criterios de causalidad con mayor certeza que aprueban la relación causa-efecto.

La variable dependiente no siempre es única, por ejemplo si aplicamos captopril a un paciente hipertenso el efecto que deseamos lograr es una modificación, pero sabemos que además del efecto deseado sobre la presión arterial, produce tos como un efecto adverso. Pero la variable de estudio sigue siendo única, aun cuando aparezcan otros efectos a partir de la intervención o manipulación.

## **4. Nivel investigativo predictivo**

### **4.1. Las variables exógenas**

Los estudios que se desarrollan en el nivel investigativo predictivo tienen como finalidad calcular la probabilidad de ocurrencia de un suceso, como, por ejemplo, una complicación de la enfermedad; también tienen la finalidad de calcular el tiempo medio en que ocurriría. Para ello, necesitamos un conjunto de características que nos permitan hacer esta predicción, estas características se denominan variables exógenas

Las variables exógenas, eran las variables independientes del nivel explicativo; pero no todas las variables independientes se convierten en variables exógenas, muchas de las variables independientes están autocorrelacionadas o son redundantes. Debemos hacer una selección de variables para construir el modelo que nos permita predecir el problema o la enfermedad que estamos estudiando, es decir, la variable de estudio.

A las variables exógenas se las conoce también como variables predictoras y antes de ingresar al modelo predictivo requieren de algunas modificaciones, en las variables dicotómicas se debe elegir a la categoría de interés, en las variables politómicas se debe construir las variables Dummy, en las variables ordinales es preciso tratarlas como si fuesen numéricas; y a las variables numéricas no se les hace ninguna modificación.

Estas modificaciones son necesarias por ejemplo en las regresiones logísticas como lineales; pero en los árboles de regresión no es necesario hacer ninguna transformación, de manera que los árboles de regresión son una buena forma de iniciar a construir un modelo predictivo mediante las técnicas de minería de datos.



## 4.2. La variable endógena

Se trata de la variable a predecir, la variable de estudio o variable endógena puede ser tanto numérica como categórica, por ejemplo, si queremos hacer la predicción de una variable categórica podemos realizar una regresión logística; y si queremos aplicar la predicción sobre una variable numérica, entonces, tendremos que realizar una regresión lineal.

Estos dos son los procedimientos más básicos para hacer predicciones en función a la ocurrencia de un determinado evento; pero no son las únicas formas de predecir. De hecho, si la variable dependiente o la variable endógena es categórica, puedes utilizar métodos de clasificación para predecir como KNN, denominado también como vecinos más cercanos.

Por otro lado, si tu variable endógena o variable de estudio es numérica puedes aplicar también los árboles de regresión, esto es una forma de predecir un resultado numérico y, por supuesto que, no son las únicas técnicas para predecir variables.

Las predicciones se ejecutan mediante ecuaciones estructurales, se construye una fórmula, o ecuación matemática, que incluye a las variables exógenas. También, es posible predecir el resultado de una variable en función de su pasado así tenemos a las series de tiempo, donde la variable exógena es el tiempo y la variable endógena es la variable de estudio.

En síntesis una variable endógena es aquella que se puede predecir en función de las variables exógenas. La variable endógena se llamó variable dependiente en el nivel explicativo; variable de supervisión en el nivel relacional, variable de interés en el nivel descriptivo, la misma que se originó en el nivel investigativo exploratorio.