Start-Tech Academy

# Dummy Variable

A Dummy variable or Indicator Variable is an artificial variable created to represent an attribute with two or more distinct categories/levels.

## Dummy Variable

### Why

- Regression analysis treats all independent (X) variables in the analysis as numerical.
- Nominal variables, or variables that describe a characteristic using two or more categories, are commonplace in regression research, but are not always useable in their categorical form.
- Dummy coding is a way of incorporating nominal variables into regression analysis

### How

- We can make a separate column, or variable, for each category.
- This new variables can take value 0 or 1 depending on the value of the categorical variable

# Dummy Variable

A Dummy variable or Indicator Variable is an artificial variable created to represent an attribute with two or more distinct categories/levels.

## Dummy Variable Example

| Student | Favorite class | Science | Math |
|---------|---------------|---------|------|
| 1 | Science | 1 | 0 |
| 2 | Science | 1 | 0 |
| 3 | English | 0 | 0 |
| 4 | Math | 0 | 1 |

**Things to keep in mind**

- The number of dummy variables necessary to represent a single attribute variable is equal to the number of levels (categories) in that variable minus one.
- We cannot code variables like science = 1, math = 2, and English = 3. As, we can see that there is no such thing as an increase in favorite class – math is not higher than science, and is not lower than language either. And even if there is increase , we cannot quantify that increase