

Light.sas: Explanation of code

Goals of code:

- Construct indicator variables “by hand”
- Demonstrate how the class statement generates indicator variables

Indicator variables by hand: data step commands

There (at least) two ways to construct indicator variables by hand. One is to use if / else to set the indicator variable (X2) to 1 or 0, depending on the value of the categorical variable (time).

The shortcut method is (`time = 'E'`). The result of `time = 'E'` is a logical value: True when `time = 'E'` and False when it is not. This can be used as an indicator variable because True is stored as a 1 and False is stored as a 0.

The interaction variable is calculated as the product of the indicator and the continuous variable. The role and interpretation of interactions will be discussed in lecture.

When you look at the printout of the data set, you see that X2 and X2a (the indicator variable constructed two different ways) have the same values. You also see that X3 is 0 for all late plants and the light value for all early plants.

The indicator and interaction variables can be used in a regression (proc reg or proc glm) just as you would any other variable.

Constructing indicator variables in proc glm:

proc glm, but not proc reg, can construct indicator variables for you. That’s what the class statement does. When a variable named in a class statement is used on the right-hand side of the model statement, SAS creates an indicator variable for each unique level of that variable.

SAS’s method of estimating the parameters is equivalent to dropping the indicator variable for the last level, which forces the “estimate” for the last level to be zero.

You may have noticed that proc glm prints the parameter estimates for a regression model but doesn’t for an ANOVA model. If there is a class statement, proc glm omits the estimates (whether or not that variable is ever used in the model). To see the estimates even when there is a class statement, add `/solution` to the model statement.

The second proc reg and the proc glm fit the same model (with an interaction of light and time). If you compare the estimates, you’ll see they are the same (and all other quantities associated with the regression are the same).