LOGISTIC REGRESSION

Logistic Regression

- Form of regression that allows the prediction of discrete variables by a mix of continuous and discrete predictors.

- Addresses the same questions that discriminant function analysis and multiple regression do but with no distributional assumptions on the predictors (the predictors do not have to be normally distributed, linearly related or have equal variance in each group)

Types of logistic regression

- ***BINARY LOGISTIC REGRESSION***
  It is used when the dependent variable is dichotomous.

MULTINOMIAL LOGISTIC REGRESSION
  It is used when the dependent or outcomes variable has more than two categories.

Binary logistic regression expression

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots\ldots b_kX_{k+E}$$

BINARY

Y = Dependent Variables

ß_0 = Constant

ß_1 = Coefficient of variable X_1

X_1 = Independent Variables

E = Error Term

Logistic Regression

In logistic regression the outcome variable is binary, and the purpose of the analysis is to assess the effects of multiple explanatory variables, which can be numeric and/or categorical, on the outcome variable.

When and Why Binary Logistic Regression?

- When the dependent variable is non parametric and we don't have homoscedasticity (variance of DV and IV not equal).

- Used when the dependent variable has only two levels. (Yes/no, male/female, taken/not taken)

- If multivariate normality is suspected.

- If we don't have linearity.

Who uses it in Plain words.

- Binary Logistic Regression can be used in the following situations.
  - A catalog company wants to increase the proportion of mailings that result in sales.
  - A doctor wants to accurately diagnose a possibly cancerous tumor.
  - A loan officer wants to know whether the next customer is likely to default.
- Using the Binary Logistic Regression procedure, the catalog company can send mailings to the people who are most likely to respond, the doctor can determine whether the tumor is more likely to be benign or malignant, and the loan officer can assess the risk of extending credit to a particular customer.

3. SAMPLE SIZE

➢ Very small samples have so much sampling errors.

➢ Very large sample size decreases the chances of errors.

➢ Logistic requires larger sample size than multiple regression.

➢ Hosmer and Lamshow recommended sample size greater than 400.

SAMPLE SIZE PER CATEGORY OF THE INDEPENDENT VARIABLE

❑ The recommended sample size for each group is at least 10 observations per estimated parameters.

ASSUMPTIONS

➢ No assumptions about the distributions of the predictor variables.

➢ Predictors do not have to be normally distributed

➢ Does not have to be linearly related.

➢ Does not have to have equal variance within each group.

➢ There should be a **minimum of 20 cases per predictor**, with a **minimum of 60 total cases**. These requirements need to be satisfied prior to doing statistical analysis with SPSS.

We will now go to SPSS for analysis.

Analyze → regression → binary logistic
Algebra 2 = gender + mosaic +
visualization test and parents education