Chapter 11 Analysis of Variance (ANOVA)

1) What is ANOVA

It is a technique measuring the strength of association and significant of the relationship between an independent variable that is discrete (number of groups larger than 2) and the dependent variable that is interval/ratio.

2) Logic of ANOVA

$\sum (X_i - \overline{X_T})^2$ *is what we called total sum of square* (SS_{Total})

$\sum (\overline{X_G} - \overline{X_T})^2 \times N_G$ *is what we called between sum of square* $(SS_{Between})$

$\sum (X_i - \overline{X_G})^2$ *is what we called within sum of square* (SS_{Within})

$SS_{Total} = SS_{Between} + SS_{Within}$

3) Computation of the sum of squares

$$\sum(X_i - \overline{X_T})^2 = \sum(\overline{X_G} - \overline{X_T})^2 \times N_G + \sum(X_i - \overline{X_G})^2$$

$$SS_{Total} = SS_{Between} + SS_{Within}$$

Because of those, I normally recommend computing $SS_{Total}\ and\ SS_{Between}$

4) Computing the df for between sum of square and within sum of square

$df_{between} = K - 1$

$df_{within} = N - K$

N: total number of cases in the sample; K: total number of groups in the independent variable

5) Computing the mean sum of square (MSS) for between and MSS for within

$$MSS_{Between} = \frac{SS_{between}}{df_{between}}$$

$$MSS_{within} = \frac{SS_{within}}{df_{within}}$$

6) Computing the f ratio

$$F_{df_{between};df_{within}} = \frac{MSS_{between}}{MSS_{within}}$$

7) Determine the p value

8) Computing the eta-square (E^2)

$$E^2 = \frac{SS_{Between}}{SS_{Total}}$$

9) Interpreting the eta-square

E^2 *is PRE* $(proportional\ Reduction\ in\ Error)$

Knowing the independent variable reduces errors in estimating the value of the dependent variable by X%.

$0 \leq E^2 \leq 1\ or\ 100\%$