# Chapter 11 Analysis of Variance (ANOVA)

1) Applications of ANOVA: when the independent variable is discrete with more than 2 groups, and the dependent variable is interval/ratio.

   ANOVA is both significant test through f-ratio and strength of association test through PRE (Proportion Reduction in Error) using eta-square.

2) ANOVA process

   <u>Step of Building blocks</u>

   Total Sum of Square ($SS_{Total}$)

$$SS_{Total} = \sum (X_i - \overline{X_T})^2$$

   Between Sum of Square

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

   Within Sum of Square

$$SS_{Within} = \sum (X_i - \overline{X_G})^2$$

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

   OR

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

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

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

<u>Step of degree of freedom</u>

$df_{between} = K - 1$ *whereas K is the total number of groups in the independent variable*

$df_{within} = N - K$ *whereas N is the total number of cases*

<u>Step of Mean sum of Square (MSS)</u>

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

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

<u>Step of F ratio computation and determine the p value</u>

$$F = \frac{MSS_{Between}}{MSS_{Within}}$$

Once you have the F ratio and degree of freedoms for between and within, we can check the F distribution table to determine the p value

Step of decision regarding the null hypothesis, and type of error committed

Reject or not to reject the null hypothesis, type I or type II error

Step of computing eta-square

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

Step of interpreting eta-square using PRE

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

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

ANOVA Table

| | Sum of Squares | df | MSS | F ratio | P level | Eta-square |
|---|---|---|---|---|---|---|
| Between | $\sum(\overline{X_G} - \overline{X_T})^2 \times N_G$ | K – 1 | $\frac{SS_{Between}}{df_{Between}}$ | $\frac{MSS_{Between}}{MSS_{Within}}$ | | $\frac{SS_{Between}}{SS_{Total}}$ |
| Within | $\sum(X_i - \overline{X_G})^2$ | N – K | $\frac{SS_{Within}}{df_{Within}}$ | | | |
| Total | $\sum(X_i - \overline{X_T})^2$ | N – 1 | | | | |