

Analysis of Covariance

- Combines linear regression and ANOVA
- Can be used to compare g treatments, after controlling for quantitative factor believed to be related to response (e.g. pre-treatment score)
- Can be used to compare regression equations among g groups (e.g. common slopes and/or intercepts)
- Model: (X quantitative, Z_1, \dots, Z_{g-1} dummy variables)

$$E(Y) = \alpha + \beta X + \beta_1 Z_1 + \dots + \beta_{g-1} Z_{g-1}$$

Tests for Additive Model

- Relation for group i ($i=1,\dots,g-1$): $E(Y)=\alpha+\beta X+\beta_i$
- Relation for group g : $E(Y)=\alpha+\beta X$
- $H_0: \beta_1=\dots=\beta_{g-1}=0$ (Controlling for covariate, no differences among treatments)

Interaction Model

- Regression slopes between Y and X are allowed to vary among groups

$$E(Y) = \alpha + \beta X + \beta_1 Z_1 + \cdots + \beta_{g-1} Z_{g-1} + \gamma_1 X Z_1 + \cdots + \gamma_{g-1} X Z_{g-1}$$

- Group i ($i=1, \dots, g-1$): $E(Y) = \alpha + \beta X + \beta_i + \gamma_i X = (\alpha + \beta_i) + (\beta + \gamma_i) X$
- Group g : $E(Y) = \alpha + \beta X$
- No interaction means common slopes: $\gamma_1 = \dots = \gamma_{g-1} = 0$

Inference in ANCOVA

- Model: $E(Y) = \alpha + \beta X + \beta_1 Z_1 + \dots + \beta_{g-1} Z_{g-1} + \gamma_1 X Z_1 + \dots + \gamma_{g-1} X Z_{g-1}$
- Construct 3 “sets” of independent variables:
 - $\{X\}$, $\{Z_1, Z_2, \dots, Z_{g-1}\}$, $\{XZ_1, \dots, XZ_{g-1}\}$
- Fit Complete model, containing all 3 sets.
 - Obtain SSE_C (or, equivalently R_C^2) and df_C
- Fit Reduced, model containing $\{X\}$, $\{Z_1, Z_2, \dots, Z_{g-1}\}$
 - Obtain SSE_R (or, equivalently R_R^2) and df_R
- $H_0: \gamma_1 = \dots = \gamma_{g-1} = 0$ (No interaction). Test Statistic:

$$F_{obs} = \frac{\left[\frac{SSE_R - SSE_C}{df_R - df_C} \right]}{\left[\frac{SSE_C}{df_C} \right]} = \frac{\left[\frac{R_C^2 - R_R^2}{df_R - df_C} \right]}{\left[\frac{1 - R_C^2}{df_C} \right]}$$

Inference in ANCOVA

- Test for Group Differences, controlling for covariate

$$E(Y) = \alpha + \beta X + \beta_1 Z_1 + \dots + \beta_{g-1} Z_{g-1}$$

- Fit Complete, model containing $\{X\}$, $\{Z_1, Z_2, \dots, Z_{g-1}\}$
 - Obtain SSE_C (or, equivalently R_C^2) and df_C
- Fit Reduced, model containing $\{X\}$
 - Obtain SSE_R (or, equivalently R_R^2) and df_R
- $H_0: \beta_1 = \dots = \beta_{g-1} = 0$ (No group differences) Test Statistic:

$$F_{obs} = \frac{\left[\frac{SSE_R - SSE_C}{df_R - df_C} \right]}{\left[\frac{SSE_C}{df_C} \right]} = \frac{\left[\frac{R_C^2 - R_R^2}{df_R - df_C} \right]}{\left[\frac{1 - R_C^2}{df_C} \right]}$$

Inference in ANCOVA

- Test for Effect of Covariate controlling for qualitative variable

$$E(Y) = \alpha + \beta X + \beta_1 Z_1 + \cdots + \beta_{g-1} Z_{g-1}$$

- $H_0: \beta=0$ (No covariate effect) Test Statistic:

$$t_{obs} = \frac{b}{\hat{\sigma}_b}$$

Adjusted Means

- Goal: Compare the g group means, after controlling for the covariate
- Unadjusted Means: $\bar{Y}_1, \dots, \bar{Y}_g$
- Adjusted Means: $\bar{Y}'_1, \dots, \bar{Y}'_g$ Obtained by evaluating regression equation at $X = \bar{X}$
- Comparing adjusted means (based on regression equation):

$$b_i = \bar{Y}'_i - \bar{Y}'_g \quad b_i - b_j = \bar{Y}'_i - \bar{Y}'_j$$

Multiple Comparisons of Adjusted Means

- Comparisons of each group with group g :

$$\hat{b}_i \pm t_{\alpha_C/2, N-g-1} \hat{\sigma}_{b_i} \quad i = 1, \dots, g-1$$

- Comparisons among the other $g-1$ groups:

$$(\hat{b}_i - \hat{b}_j) \pm t_{\alpha_C/2, N-g-1} \sqrt{\hat{\sigma}_{b_i}^2 + \hat{\sigma}_{b_j}^2 - 2COV(b_i, b_j)}$$

- Variances and covariances are obtained from computer software packages (SPSS, SAS)