2/2/2010 36-402/608 ADA-II H. Seltman

Handout #7: One-way ANOVA

1. Analysis of Variance is a "workhorse" of data analysis, rather than an advanced technique.

2. Concept of the F test for equal group means

   (a) Focuses on categorical explanatory variables (with Normal outcomes)

   (b) Uses variance-like-quantities as a tool to investigate mean differences.

   (c) Decompose deviations from the grand mean into meaningful components that are orthogonal. This leads to SS (sums of squared deviations) components that meaningfully add up to the total SS.

   (d) $A + B = C \Rightarrow A^2 + B^2 = C^2$ only when A and B are orthogonal, which corresponds to statistically independent.

   (e) One way ANOVA with I groups each with n_i observations $Y_{ij},\ j \in \{1, \ldots, n_i\}$:

$$Y_{ij} - \bar{Y} = \left(\bar{Y}_i - \bar{Y}\right) + \left(Y_{ij} - \bar{Y}_i\right)$$

$$\sum_{ij} \left(Y_{ij} - \bar{Y}\right) = \sum_{ij} \left[\left(\bar{Y}_i - \bar{Y}\right) + \left(Y_{ij} - \bar{Y}_i\right)\right] = 0$$

$$\sum_{ij} \left(Y_{ij} - \bar{Y}\right)^2 = \sum_{ij} \left(\bar{Y}_i - \bar{Y}\right)^2 + \sum_{ij} \left(Y_{ij} - \bar{Y}_i\right)^2 + 0$$

$$\text{SS}_{\text{Total}} = \text{SS}_{\text{Between Groups}} + \text{SS}_{\text{Within Groups}}$$

   (f) Definition: $MS = SS/df$, applies to $s_Y^2 = SS_T/(n-1)$.

   (g) Well known fact: For $Y_i \sim N(\mu, \sigma^2)$, $Z_i = \frac{Y_i - \mu_i}{\sigma}$ is standard normal. If $Z_1, \ldots, Z_p$ are independent standard normal variables, then $\sum_{i=1}^{p} Z^2$ is a chi-square random variable with p df.

   (h) Intuitively similar fact: If a sum of squares of normal deviates (with variance σ^2) has d degrees of freedom, then with SS/σ^2 has a distribution that is chi-squared with d degrees of freedom.

   (i) Well known fact: If $U \sim \chi_u^2$ and $V \sim \chi_v^2$ are independent, then $\frac{U/u}{V/v} \sim F_{u,v}$.

   (j) Conclusion: $\frac{SS_B/df_B}{SS_W/df_W} = MS_B/MS_W \sim F_{df_B, df_W}$ if both mean squares have the same variance. MS_W has variance σ_2, but MS_B has that variance only if $H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$, otherwise it has a larger variance.

3. Concept of the planned test of contrasts:

   (a) Let $C_1, ..., C_I$ be a set of coefficients adding to 0.

   (b) C defines a "contrast null hypothesis": $\gamma = C_1\mu_1 + \cdots + C_I\mu_i = 0$. E.g., with I=4 and C=(1,0,-1/2,-1/2) this is a test of whether the mean of treatment 1 differs from the average of the means of treatments 3 and 4.

   (c) The obvious estimate of γ is $G = C_1\bar{Y}_1 + \cdots + C_I\bar{Y}_I$.

   (d) If σ^2 is known, the variance pf $a\bar{Y}_1$ is $a^2\frac{\sigma^2}{n_1}$. The variance of G is $\left(\sum_{i=1}^{I} C_i^2 \frac{\sigma^2}{n_i}\right)$.

   (e) If σ^2 is unknown, the best estimate of G's variance is $s_p^2 \left(\sum_{i=1}^{I} C_i^2/n_i\right)$

   (f) s_p^2 is MS_W in ANOVA or the "residual standard error" in regression.

   (g) As usual $\text{SE}(G) = \sqrt{\text{Var}(G)}$.

   (h) Test $H_0 : \gamma = 0$ using $T = \frac{G}{\text{SE}(G)}$ with df_W degrees of freedom.

4. Breakout and Discussion