

Stat 218 - Day 28
Comparing Several Means: Analysis of Variance

Today we return to analyzing quantitative variables. Several weeks ago we studied two-sample t -tests for comparing means between two groups. This week we'll extend that procedure to comparing means between *several* groups. This resulting procedure is called **analysis of variance**, abbreviated **ANOVA**.

Example: Discrimination against handicapped

In a study of whether physical handicaps affect perceptions of employment qualifications, researchers showed videotapes of an employer interviewing an applicant, but both men were actors following a set script. In the videotapes the “applicant” appeared with different handicaps (none, amputee, crutches, hearing, wheelchair). A group of seventy students were randomly assigned to watch a videotape and evaluate the applicant’s qualifications on a ten-point scale. The study aimed to assess whether the data provide evidence that the mean qualification scores differ significantly among the various handicaps presented.

(a) Is this an observational study or an experiment? Explain.

(b) Identify the explanatory and the response variable. For each one, indicate whether it is quantitative or categorical.

ANOVA applies to situations like this where the explanatory variable is categorical and the response variable is quantitative. If the categorical explanatory variable was also binary, then we could use a two-sample t -test to analyze the data. But in this situation we have five categories to compare, and we need a test that will compare all five *simultaneously*. Before we get to the test, though, we start with graphical and numerical summaries of the data.

(c) Examine dotplots and boxplots of the qualification ratings among the five groups (HandicapApply.mtw). Then examine summary statistics. Do the groups appear to differ with regard to qualification ratings? Explain.

Now some notation:

- I = number of groups
- μ_i = population mean for group i
- n_i = sample size for group i
- $\bar{y}_{i\cdot}$ = sample mean for group i
- s_i = sample standard deviation for group i
- n^* = grand (overall) sample size
- $\bar{y}_{\cdot\cdot}$ = grand (overall) mean for all groups combined
- y_{ij} = j^{th} observation in group i

We want to test the null hypothesis that all population means are equal ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$). The key idea of ANOVA is to compare variability *between* groups to variability *within* groups. We'll measure this variability through **sums of squares** and **mean squares**.

- Total sum of squares: $SS(\text{total}) = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2$
- Total degrees of freedom: $df(\text{total}) = n^* - 1$

(d) Calculate $SS(\text{total})$ for the handicap study.

- Within-group sum of squares: $SS(\text{within}) = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^I (n_i - 1)s_i^2$
- Within-group degrees of freedom: $df(\text{within}) = n^* - I$
- Within-group mean square: $MS(\text{within}) = SS(\text{within}) / df(\text{within})$

(e) Calculate $SS(\text{within})$, $df(\text{within})$, and $MS(\text{within})$ for the handicap study.

ANOVA combines information on variability within all of the groups to create a pooled estimate of the standard deviation within a group:

- $s_{pooled} = \sqrt{MS(\text{within})}$

(f) Calculate the pooled standard deviation for the handicap study.

- Between-group sum of squares: $SS(\text{between}) = \sum_{i=1}^I n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$
- Between-group degrees of freedom: $df(\text{between}) = I - 1$
- Between-group mean square: $MS(\text{between}) = SS(\text{between}) / df(\text{between})$

(g) Calculate $SS(\text{between})$, $df(\text{between})$, and $MS(\text{between})$ for the handicap study.

- $SS(\text{total}) = SS(\text{between}) + SS(\text{within})$

(h) Verify that this relationship holds for the handicap study.

To organize all of these calculations, we construct an **ANOVA table**.

(i) Create an ANOVA table for the handicap study. Then use Minitab to verify this table (Stat> Anova> Oneway).

- Test statistic: $F_s = MS(\text{between}) / MS(\text{within})$

(j) *Key question*: Do large or small values of F_s provide evidence against the null hypothesis that all population means are equal? Explain.

(k) Calculate the value of the test statistic F_s for the handicap study.

Is this value large? To answer this, we need to compare it to a reference distribution in order to calculate the P -value.

- P -value: obtained from F -distribution (Table 10) with numerator $df = df(\text{between})$ and denominator $df = df(\text{within})$

(k) Determine the P -value as accurately as possible.

(l) Would you reject the null hypothesis at the $\alpha = .05$ significance level? Explain.

(m) Summarize the conclusion that you would draw from this study.

(n) What would change in the ANOVA table if the sample means were further apart? How would the P -value change? How would your conclusion change? Explain.

(o) What would change in the ANOVA table if the qualification ratings in each group were further apart? How would the P -value change? How would your conclusion change? Explain.