Chapter 13
Analysis of Variance (ANOVA)

ANOVA is an extension of what you have learned in Chapter 9. ANOVA techniques compare the means of **several groups** (two or more populations) using an independent sample from each group.

Identifying groups as a categorical variable, we have a quantitative variable (the variable of interest) that is explained by a categorical variable (the groups).

In ANOVA, the categorical variables that identify the groups are called the **factors.** When there is only one factor the technique is called **one–way ANOVA**.

13.1 One–way ANOVA

Suppose there is one factor with g levels (groups or populations, called **treatments** in ANOVA). In each treatment group we measure a quantitative response variable and compare the means of the response in each treatment group.

One of the hypotheses of interest is the equality of the population means, i.e., **Ho: $\mu_1 = \mu_2 = \dots = \mu_g$.** If Ho is not true, what can we say about the relation between the means?

Ha: At least two of the population means are unequal.

Equivalently we may specify these hypotheses as

$$\text{Ho: } \mu_1 = \mu_2 = \ldots = \mu_g = \mu \text{ vs.}$$
$$\text{Ha: At least one of } \mu_i \neq \mu.$$

[See Figure 13.2 for a graphical example]

Analysis of variance:

To test if there are significant differences between the population means the **test statistic** is

$$F = \frac{\textit{Variability between groups}}{\textit{Variability within groups}} = \frac{s^2_{Between}}{s^2_{Within}}$$

- If **thereIS a significant difference** between group means, then variability between groups >> variability within groups, i.e., F will be very large.

- If **there is NO significant difference** between group means, then variability between groups will be approximately equal to the variability within groups, i.e. $F \approx 1$

ANOVA TESTS:

1. <u>Assumptions</u>
   a) Random samples
   b) Normal populations (i.e., distribution of the response variable is normal in each group). Critical for small samples, not too important for large samples.
   c) Equal variances for all treatment groups (populations). If sample sizes are equal then this assumption is not crucial. Otherwise check if **two times the smallest sample standard deviation** is more than the largest sample standard deviation.
2. <u>Hypotheses:</u>
   Null hypothesis: Ho: $\mu_1 = \mu_2 = \ldots = \mu_g = \mu$
   Alternative hypothesis: Ha: At least one $\mu_i \neq \mu$
3. <u>Test statistic</u>: F = MSB / MSW
4. <u>The p-value</u>: Right tail probability
   Compare with values from F-tables at the intersection of $df_1 = g - 1$ and $df_2 = N - g$, where N = sum of all samples = Total number of observations and g = number of groups.
5. <u>Decision Rule:</u> Reject Ho if p-value $\leq \alpha$ always.
6. <u>Conclusion:</u> State decision in layman's language