Chapter 12
Analysis of Variance

12-1 Overview

12-2 One-Way ANOVA

Key Concept

This section introduces the method of one-way analysis of variance, which is used for tests of hypotheses that three or more population means are all equal.

Overview

- Analysis of variance (ANOVA) is a method for testing the hypothesis that three or more population means are equal.
- For example:
  \[ H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k \]
  \[ H_1: \text{At least one mean is different} \]

ANOVA Methods

Require the \( F \)-Distribution

1. The \( F \)-distribution is not symmetric; it is skewed to the right.
2. The values of \( F \) can be 0 or positive; they cannot be negative.
3. There is a different \( F \)-distribution for each pair of degrees of freedom for the numerator and denominator.

Critical values of \( F \) are given in Table A-5

One-Way ANOVA

An Approach to Understanding ANOVA

1. Understand that a small \( P \)-value (such as 0.05 or less) leads to rejection of the null hypothesis of equal means. With a large \( P \)-value (such as greater than 0.05), fail to reject the null hypothesis of equal means.
2. Develop an understanding of the underlying rationale by studying the examples in this section.
One-Way ANOVA
An Approach to Understanding ANOVA

3. Become acquainted with the nature of the SS (sum of squares) and MS (mean square) values and their role in determining the $F$ test statistic, but use statistical software packages or a calculator for finding those values.

Definition
Treatment (or factor)
A treatment (or factor) is a property or characteristic that allows us to distinguish the different populations from one another.

Use computer software or TI-83/84 for ANOVA calculations if possible.

One-Way ANOVA
Requirements
1. The populations have approximately normal distributions.
2. The populations have the same variance $\sigma^2$ (or standard deviation $\sigma$).
3. The samples are simple random samples.
4. The samples are independent of each other.
5. The different samples are from populations that are categorized in only one way.

Procedure for testing
$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots$
1. Use STATDISK, Minitab, Excel, or a TI-83/84 Calculator to obtain results.
2. Identify the $P$-value from the display.
3. Form a conclusion based on these criteria:
   - If $P$-value $\leq \alpha$, reject the null hypothesis of equal means.
   - If $P$-value $> \alpha$, fail to reject the null hypothesis of equal means.

Example: Weights of Poplar Trees
Do the samples come from populations with different means?

Caution when interpreting ANOVA results:
When we conclude that there is sufficient evidence to reject the claim of equal population means, we cannot conclude from ANOVA that any particular mean is different from the others. There are several other tests that can be used to identify the specific means that are different, and those procedures are called multiple comparison procedures, and they are discussed later in this section.
Example: Weights of Poplar Trees

Do the samples come from populations with different means?

$H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1$: At least one of the means is different from the others.

For a significance level of $\alpha = 0.05$, use STATDISK, Minitab, Excel, or a TI-83/84 calculator to test the claim that the four samples come from populations with means that are not all the same.

Example: Weights of Poplar Trees

Do the samples come from populations with different means?

$H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1$: At least one of the means is different from the others.

The displays all show a $P$-value of approximately 0.007. Because the $P$-value is less than the significance level of $\alpha = 0.05$, we reject the null hypothesis of equal means. There is sufficient evidence to support the claim that the four population means are not all the same. We conclude that those weights come from populations having means that are not all the same.

ANOVA Fundamental Concepts

Estimate the common value of $\sigma^2$:

1. The variance between samples (also called variation due to treatment) is an estimate of the common population variance $\sigma^2$ that is based on the variability among the sample means.

2. The variance within samples (also called variation due to error) is an estimate of the common population variance $\sigma^2$ based on the sample variances.

Test Statistic for One-Way ANOVA

$F = \frac{\text{variance between samples}}{\text{variance within samples}}$

An excessively large $F$ test statistic is evidence against equal population means.
Relationships Between the \( F \) Test Statistic and \( P \)-Value

Figure 12-2

Calculations with Equal Sample Sizes

\( \star \) Variance between samples = \( n s_x^2 \)

where \( s_x^2 \) = variance of sample means

\( \star \) Variance within samples = \( s_p^2 \)

where \( s_p^2 \) = pooled variance (or the mean of the sample variances)

Example: Sample Calculations

Use Table 12-2 to calculate the variance between samples, variance within samples, and the \( F \) test statistic.

1. Find the variance between samples = \( n s_x^2 \).

For the means 5.5, 6.0 & 6.0, the sample variance is

\[ s_x^2 = 0.0833 \]

\[ n s_x^2 = 4 \times 0.0833 = 0.3332 \]

2. Estimate the variance within samples by calculating the mean of the sample variances.

\[ s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333 \]

3. Evaluate the \( F \) test statistic

\[ F = \frac{\text{variance between samples}}{\text{variance within samples}} \]

\[ F = \frac{0.3332}{2.3333} = 0.1428 \]

Critical Value of \( F \)

\( \star \) Right-tailed test

\( \star \) Degree of freedom with \( k \) samples of the same size \( n \)

numerator \( df = k - 1 \)

denominator \( df = k(n - 1) \)
Calculations with Unequal Sample Sizes

\[ F = \frac{\text{variance within samples}}{\text{variance between samples}} = \frac{\frac{\sum n_i (x_i - \bar{x})^2}{k-1}}{\frac{\sum (n_i-1)x_i^2}{\sum (n_i-1)}} \]

where \( \bar{x} \) = mean of all sample scores combined

\( k \) = number of population means being compared

\( n_i \) = number of values in the \( i \)th sample

\( x_i \) = mean of values in the \( i \)th sample

\( s_i^2 \) = variance of values in the \( i \)th sample

Key Components of the ANOVA Method

**SS(total)**, or total sum of squares, is a measure of the total variation (around \( \bar{x} \)) in all the sample data combined.

Formula 12-1

\[ SS(\text{total}) = \sum (x - \bar{x})^2 \]

Key Components of the ANOVA Method

**SS(treatment)**, also referred to as SS(factor) or SS(between groups) or SS(between samples), is a measure of the variation between the sample means.

Formula 12-2

\[ SS(\text{treatment}) = n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \ldots + n_k(x_k - \bar{x})^2 = \sum n_i(x_i - \bar{x})^2 \]

Key Components of the ANOVA Method

Given the previous expressions for SS(total), SS(treatment), and SS(error), the following relationship will always hold.

Formula 12-4

\[ SS(\text{total}) = SS(\text{treatment}) + SS(\text{error}) \]

Mean Squares (MS)

**MS(treatment)** is a mean square for treatment, obtained as follows:

Formula 12-5

\[ MS(\text{treatment}) = \frac{SS(\text{treatment})}{k-1} \]

**MS(error)** is a mean square for error, obtained as follows:

Formula 12-6

\[ MS(\text{error}) = \frac{SS(\text{error})}{N-k} \]
**Mean Squares (MS)**

$MS_{(total)}$ is a mean square for the total variation, obtained as follows:

**Formula 12-7**

$$MS_{(total)} = \frac{SS_{(total)}}{N - 1}$$

**Test Statistic for ANOVA with Unequal Sample Sizes**

**Formula 12-8**

$$F = \frac{MS \text{ (treatment)}}{MS \text{ (error)}}$$

- Numerator $df = k - 1$
- Denominator $df = N - k$

**Example: Weights of Poplar Trees**

Table 12-3 has a format often used in computer displays.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares (SS)</th>
<th>Degrees of Freedom (MS)</th>
<th>F Test Statistic</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>4.6824</td>
<td>3</td>
<td>1.5608</td>
<td>0.0073</td>
</tr>
<tr>
<td>Error</td>
<td>4.3572</td>
<td>16</td>
<td>0.2723</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9.0397</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Identifying Means That Are Different**

Informal procedures to identify the specific means that are different

1. Use the same scale for constructing boxplots of the data sets to see if one or more of the data sets are very different from the others.
2. Construct confidence interval estimates of the means from the data sets, then compare those confidence intervals to see if one or more of them do not overlap with the others.

**Identifying Means That Are Different**

After conducting an analysis of variance test, we might conclude that there is sufficient evidence to reject a claim of equal population means, but we cannot conclude from ANOVA that any particular mean is different from the others.

**Bonferroni Multiple Comparison Test**

Step 1. Do a separate $t$ test for each pair of samples, but make the adjustments described in the following steps.

Step 2. For an estimate of the variance $\sigma^2$ that is common to all of the involved populations, use the value of $MS(error)$. 
Bonferroni Multiple Comparison Test

Step 2 (cont.) Using the value of MS(error), calculate the value of the test statistic, as shown below. (This example shows the comparison for Sample 1 and Sample 4.)

\[
t = \frac{\bar{x}_1 - \bar{x}_4}{\sqrt{MS(error) \left(\frac{1}{n_1} + \frac{1}{n_4}\right)}}
\]

Bonferroni Multiple Comparison Test

Step 2 (cont.) Change the subscripts and use another pair of samples until all of the different possible pairs of samples have been tested.

Bonferroni Multiple Comparison Test

Step 3. After calculating the value of the test statistic \(t\) for a particular pair of samples, find either the critical \(t\) value or the \(P\)-value, but make the following adjustment:

Step 3 (cont.) \(P\)-value: Use the test statistic \(t\) with \(df = N-k\), where \(N\) is the total number of sample values and \(k\) is the number of samples. Find the \(P\)-value the usual way, but adjust the \(P\)-value by multiplying it by the number of different possible pairings of two samples.

(For example, with four samples, there are six different possible pairings, so adjust the \(P\)-value by multiplying it by 6.)

Bonferroni Multiple Comparison Test

Step 3 (cont.) Critical value: When finding the critical value, adjust the significance level \(\alpha\) by dividing it by the number of different possible pairings of two samples.

(For example, with four samples, there are six different possible pairings, so adjust the \(P\)-value by dividing it by 6.)

Example: Weights of Poplar Trees Using the Bonferroni Test

Using the data in Table 12-1, we concluded that there is sufficient evidence to warrant rejection of the claim of equal means.

The Bonferroni test requires a separate \(t\) test for each different possible pair of samples.

\[
\begin{align*}
H_0: \mu_1 &= \mu_2 & H_0: \mu_1 &= \mu_3 & H_0: \mu_1 &= \mu_4 \\
H_0: \mu_2 &= \mu_1 & H_0: \mu_2 &= \mu_3 & H_0: \mu_2 &= \mu_4 \\
H_0: \mu_3 &= \mu_1 & H_0: \mu_3 &= \mu_2 & H_0: \mu_3 &= \mu_4 \\
H_0: \mu_4 &= \mu_1 & H_0: \mu_4 &= \mu_2 & H_0: \mu_4 &= \mu_3 \\
\end{align*}
\]

We begin with testing \(H_0: \mu_1 = \mu_4\).

From Table 12-1: \(\bar{x}_1 = 0.184, \ n_1 = 5, \ \bar{x}_4 = 1.334, \ n_4 = 5\)

\[
t = \frac{\bar{x}_1 - \bar{x}_4}{\sqrt{MS(error) \left(\frac{1}{n_1} + \frac{1}{n_4}\right)}} = \frac{0.184 - 1.334}{\sqrt{0.2723275 \left(\frac{1}{5} + \frac{1}{5}\right)}} = -3.484
\]
**Example: Weights of Poplar Trees**

**Using the Bonferroni Test**

Test statistic $= -3.484$

$df = N - k = 20 - 4 = 16$

Two-tailed $P$-value is 0.003065, but adjust it by multiplying by 6 (the number of different possible pairs of samples) to get $P$-value $= 0.01839$.

Because the adjusted $P$-value is less than $\alpha = 0.05$, reject the null hypothesis.

It appears that Samples 1 and 4 have significantly different means.

---

**SPSS Bonferroni Results**

<table>
<thead>
<tr>
<th>Samples</th>
<th>Mean Difference</th>
<th>t</th>
<th>df</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 4</td>
<td>-6.400</td>
<td>-3.484</td>
<td>16</td>
<td>0.003065</td>
<td>(-11.460, -1.440)</td>
</tr>
<tr>
<td>2 and 3</td>
<td>0.450</td>
<td>0.058</td>
<td>16</td>
<td>0.9564</td>
<td>(-0.342, 1.242)</td>
</tr>
<tr>
<td>2 and 4</td>
<td>-1.020</td>
<td>-0.895</td>
<td>16</td>
<td>0.3866</td>
<td>(-2.560, 0.440)</td>
</tr>
<tr>
<td>3 and 4</td>
<td>-1.520</td>
<td>-1.602</td>
<td>16</td>
<td>0.1191</td>
<td>(-3.560, 0.520)</td>
</tr>
</tbody>
</table>

*This mean difference is significant at the 0.05 level.*