12.1 The F distribution

We already covered this topic in Ch. 10
12.2 One-Way Analysis of Variance

- The analysis of variance procedure is used to test the null hypothesis that the means of three or more populations are the same against the alternative hypothesis that not all population means are the same.
- The analysis of variance procedure can be used to compare two population means. However, the procedures learned in Chapter 10 are more efficient.
- For example if teachers at a school have devised three different methods to teach arithmetic and want to find out if these three methods produce different mean scores.
- Let \( \mu_1, \mu_2, \) and \( \mu_3 \) be the mean scores of all students who are taught by Methods I, II, and III, respectively.

To test if the three teaching methods produce different means, we test the null hypothesis against the alternative hypothesis

\[
H_0 : \mu_1 = \mu_2 = \mu_3
\]

Against

\[
H_1 : \text{Not all the three population means are equal}
\]

To test such a hypothesis using method learned in CH. 10 is to test the three hypotheses \( H0: \mu_1 = \mu_2, \) \( H0: \mu_1 = \mu_3, \) and \( H0: \mu_2 = \mu_3 \)

Despite the time consumption, it is not recommended procedure due to the high type I error associated with it when it is used for more than 2 populations.
12.2 One-Way Analysis of Variance

- The ANOVA, short for analysis of variance, is used to compare three or more population means in a single test.
- This section discusses the one-way ANOVA procedure to make tests comparing the means of several populations.
- By using a one-way ANOVA test, we analyze only one factor or variable.
- In the example of testing for the equality of mean arithmetic scores of students taught by each of the three different methods, we are considering only one factor, which is the effect of different teaching methods on the scores of students.

Sometimes we may analyze the effects of two factors.
- If different teachers teach arithmetic using these three methods, we can analyze the effects of teachers and teaching methods on the scores of students. This is done by using a two-way ANOVA.
- The procedure under discussion in this chapter is called the analysis of variance because the test is based on the analysis of variation in the data obtained from different samples.
12.2 One-Way Analysis of Variance

The application of one-way ANOVA requires that the following assumptions hold true.

1. The populations from which the samples are drawn are (approximately) normally distributed.
2. The populations from which the samples are drawn have the same variance (or standard deviation).
3. The samples drawn from different populations are random and independent.

The ANOVA test is applied by calculating two estimates of the variance, $\sigma^2$, of population distributions: the variance between samples and the variance within samples.

The variance between samples is also called the mean square between samples or $MSB$.

The variance within samples is also called the mean square within samples or $MSW$.

The variance between samples, $MSB$, gives an estimate of $\sigma^2$ based on the variation among the means of samples taken from different populations.

For the example of three teaching methods, $MSB$ will be based on the values of the mean scores of three samples of students taught by three different methods.
12.2 One-Way Analysis of Variance

- If the means of all populations under consideration are equal, the means of the respective samples will still be different but the variation among them is expected to be small.
- Consequently, the value of MSB is expected to be small.
- However, if the means of populations under consideration are not all equal, the variation among the means of respective samples is expected to be large.
- Consequently, the value of MSB is expected to be large.

The variance within samples, MSW, gives an estimate of $\sigma^2$ based on the variation within the data of different samples.

- For the example of three teaching methods, MSW will be based on the scores of individual students included in the three samples taken from three populations.
- The concept of MSW is similar to the concept of the pooled standard deviation, $sp$, for two samples discussed earlier in CH 10.
- The one-way ANOVA test is always right-tailed with the rejection region in the right tail of the $F$ distribution curve.
Chapter 11: Chi-Square test

12.2 One-Way Analysis of Variance

12.2.1 Calculating the Value of the Test Statistic

The value of the test statistic $F$ for a test of hypothesis using ANOVA is given by the ratio of two variances, the variance between samples (MSB) and the variance within samples (MSW).

$$F = \frac{\text{Variance between samples}}{\text{Variance within samples}} = \frac{\text{MSB}}{\text{MSW}}$$

The between-samples sum of squares, denoted by SSB, is calculated as

$$SSB = \left( \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \cdots \right) - \frac{(\sum x)^2}{n}$$

where

- $x$ = the score of a student
- $k$ = the number of different samples (or treatments)
- $n_i$ = the size of sample $i$
- $T_i$ = the sum of the values in sample $i$
- $n = \text{the number of values in all samples} = n_1 + n_2 + n_3 + \cdots$
- $\Sigma x$ = the sum of the values in all samples = $T_1 + T_2 + T_3 + \cdots$
- $\Sigma x^2$ = the sum of the squares of the values in all samples
12.2 One-Way Analysis of Variance

12.2.1 Calculating the Value of the Test Statistic

- The values of MSB and MSW are calculated as

\[ MSB = \frac{SSB}{k-1} \quad \text{and} \quad MSW = \frac{SSW}{n-k} \]

where \( k - 1 \) and \( n - k \) are, respectively, the \( df \) for the numerator and the \( df \) for the denominator for the \( F \) distribution. Remember, \( k \) is the number of different samples.

12.2.2 One-Way ANOVA Test

Example: Fifteen fourth-grade students were randomly assigned to three groups to experiment with three different methods of teaching arithmetic. At the end of the semester, the same test was given to all 15 students. The table gives the scores of students in the three groups. At the 1% significance level, can we reject the null hypothesis that the mean arithmetic score of all fourth-grade students taught by each of these three methods is the same? Assume that all the assumptions required to apply the one-way ANOVA procedure hold true.

<table>
<thead>
<tr>
<th>Method I</th>
<th>Method II</th>
<th>Method III</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>55</td>
<td>84</td>
</tr>
<tr>
<td>73</td>
<td>85</td>
<td>68</td>
</tr>
<tr>
<td>51</td>
<td>70</td>
<td>95</td>
</tr>
<tr>
<td>65</td>
<td>69</td>
<td>74</td>
</tr>
<tr>
<td>87</td>
<td>90</td>
<td>67</td>
</tr>
</tbody>
</table>
12.2 One-Way Analysis of Variance

Step 1. State the null and alternative hypotheses.
Let $\mu_1$, $\mu_2$, and $\mu_3$ be the mean arithmetic scores of all fourth-grade students who are taught, respectively, by Methods I, II, and III. The null and alternative hypotheses are

$H_0: \mu_1 = \mu_2 = \mu_3$  \text{Vs.}  \ H_1: \text{Not all the means are equal}$

Step 2. Select the distribution to use.
Because we are comparing the means for three normally distributed populations, we use the $F$ distribution to make this test.

Step 3. Determine the rejection and nonrejection regions.
The significance level is .01. Because a one-way ANOVA test is always right-tailed, the area in the right tail of the $F$ distribution curve is .01

Numerator degrees of freedom = $k - 1 = 3 - 1 = 2$
Denominator degrees of freedom = $n - k = 15 - 3 = 2$
12.2 One-Way Analysis of Variance

Step 4. Calculate the value of the test statistic.

\[
T_1 = 48 + 73 + 51 + 65 + 87 = 324 \\
T_2 = 55 + 85 + 70 + 69 + 90 = 369 \\
T_3 = 84 + 68 + 95 + 74 + 67 = 388 \\
\sum x = T_1 + T_2 + T_3 = 1081 \\
n = 5 + 5 + 5 = 15 \\
\sum x^2 = (48)^2 + (73)^2 + (51)^2 + \ldots + (74)^2 + (67)^2 = 80,709 \\
SSB = \left(\frac{(324)^2}{5} + \frac{(369)^2}{5} + \frac{(388)^2}{5}\right) - \frac{(1081)^2}{15} = 432.1333 \\
SSW = 80,709 - \left(\frac{(324)^2}{5} + \frac{(369)^2}{5} + \frac{(388)^2}{5}\right) = 2372.8000 \\
\]

\[
\begin{array}{c|c|c|c}
\text{Method I} & \text{Method II} & \text{Method III} \\
\hline
48 & 55 & 84 \\
73 & 85 & 68 \\
51 & 70 & 95 \\
65 & 69 & 74 \\
87 & 90 & 67 \\
\end{array}
\]

\[
F = \frac{MSB}{MSW} = \frac{216.0667}{197.7333} = \]

\[
\frac{T_2 - T_1}{5} \quad \frac{T_2 - T_3}{5} \quad \frac{T_3 - T_1}{5} \\
55 \quad 68 \quad 94 \\
70 \quad 74 \quad 67 \\
84 \quad 87 \quad 90 \\
\]
### Chapter 11: Chi-Square test

#### 12.2 One-Way Analysis of Variance

- For convenience, all these calculations are often recorded in a table called the ANOVA table.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>Value of the Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>2</td>
<td>432.1333</td>
<td>216.0667</td>
<td>( F = \frac{216.0667}{197.7333} = 1.09 )</td>
</tr>
<tr>
<td>Within</td>
<td>12</td>
<td>2372.8000</td>
<td>197.7333</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>2804.9333</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Step 5. Make a decision.**

- The test statistic \( F = 1.09 \) is less than the \( F \) critical (6.93).
- We fail to reject the null hypothesis and conclude that the means of the three populations are equal.

---

Example: From time to time, unknown to its employees, the research department at Post Bank observes various employees for their work productivity. Recently this department wanted to check whether the four tellers at a branch of this bank serve, on average, the same number of customers per hour. The research manager observed each of the four tellers for a certain number of hours. The following table gives the number of customers served by the four tellers during each of the observed hours.

At the 5% significance level, test the null hypothesis that the mean number of customers served per hour by each of these four tellers is the same. Assume that all the assumptions required to apply the one-way ANOVA procedure hold true.