

# Comparing The Means of Many Independent Samples

In this chapter we study analysis of variance (ANOVA). In Chapter 7 we considered the comparison of two independent samples with respect to a quantitative variable  $Y$ . The classical techniques for comparing the two sample means  $\bar{Y}_1$  and  $\bar{Y}_2$  are the test and the confidence interval based on Student's  $t$  distribution. In this chapter we consider the comparison of the means of  $I$  independent samples, where  $I$  may be greater than 2.

## Example: sweet corn

When growing sweet corn, can organic methods be used successfully to control harmful insects and limit their effect on the corn? In a study of this question researchers compared the weights of ears of corn under five conditions in an experiment in which sweet corn was grown using organic methods. The treatments were

- Treatment 1: Nematodes
- Treatment 2: Wasps
- Treatment 3: Nematodes and wasps
- Treatment 4: Bacteria
- Treatment 5: Control

Ears of corn were randomly sampled from each plot and weighed. The results are given in the table and figure below.

Table 11.1.1 Weights (ounces) of ears of sweet corn					
	Treatment				
	1	2	3	4	5
	16.5	11.0	8.5	16.0	13.0
	15.0	15.0	13.0	14.5	10.5
	11.5	9.0	12.0	15.0	11.0
	12.0	9.0	10.0	9.0	10.0
	12.5	11.5	12.5	10.5	14.0
	9.0	11.0	8.5	14.0	12.0
	16.0	9.0	9.5	12.5	11.0
	6.5	10.0	7.0	9.0	9.5
	8.0	9.0	10.5	9.0	18.5
	14.5	8.0	10.5	9.0	17.0
	7.0	8.0	13.0	6.5	10.0
	10.5	5.0	9.0	8.5	11.0
Mean	11.58	9.63	10.33	11.13	12.29
SD	3.47	2.42	1.96	3.12	2.87
$n$	12	12	12	12	12

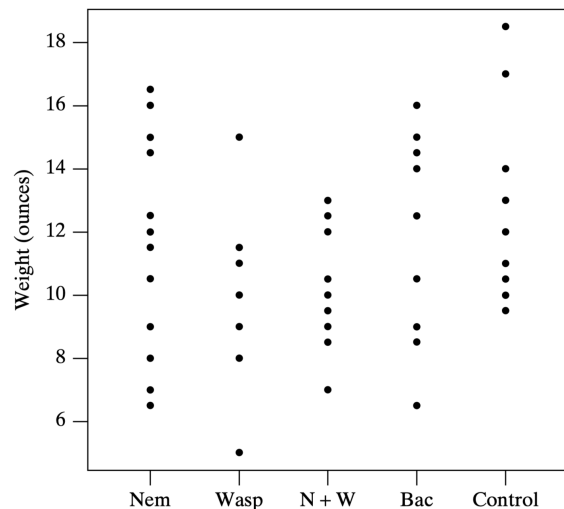


Figure 11.1.1 Weights of ears of corn receiving five different treatments

The classical method of analyzing data from  $I$  independent samples is called an **analysis of variance**, or **ANOVA**. In applying analysis of variance, the data are regarded as **random samples** from  $I$  populations. We denote the means of these populations as  $\mu_1, \mu_2, \dots, \mu_I$  and the standard deviations as  $\sigma_1, \sigma_2, \dots, \sigma_I$ . We test a null hypothesis of equality among all  $I$  population means,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_I.$$

## Why not repeated $t$ tests?

- It is natural to wonder why the comparison of the means of  $I$  samples requires any new methods. For instance, why not just use a two-sample  $t$  test on each pair of samples?
- The most serious difficulty with a naive "repeated  $t$  tests" procedure concerns Type I error: *The probability of false rejection of a null hypothesis may be much higher than it appears to be.*
- For instance, suppose  $I = 4$  and consider the null hypothesis that all four population means are equal ( $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ) versus the alternative hypothesis that the four means are not all equal. Among four means there are six possible pairs to compare.
- Let's consider the risk of a Type I error for testing our primary null hypothesis that all four means are equal by conducting six separate two-sample  $t$  tests. If **any** of the six  $t$  tests finds a significant difference between a pair of means, we would reject our primary null hypothesis that all four means are equal. A Type I error would occur if **any** of the six  $t$  tests found a significant difference between a pair of means when in fact all four means are equal.

The table below displays the overall risk of Type I error. It is clear that the researcher who uses repeated  $t$  tests is highly vulnerable to Type I error unless  $I$  is quite small. The difficulties illustrated by the table below are due to **multiple comparisons**. That is, many comparisons on the same set of data. These difficulties can be reduced when the comparison of several groups is approached through ANOVA.

<b>Table 11.1.2 Overall risk of type I error in using repeated <math>t</math> tests at <math>\alpha = 0.05</math></b>	
$I$	Overall risk
2	0.05
3	0.12
4	0.20
6	0.37
8	0.51
10	0.63

## The Basic One-Way Analysis of Variance

The ANOVA model presented previously that compares the means of three or more groups is called a one-way ANOVA. The term "one-way" refers to the fact that there is one variable that defines the groups or treatments (e.g., in the sweet corn example the treatments were based on the type of harmful insect/bacteria).

## Notation

- The  $j$ th observation in group  $i$ :  $Y_{ij}$
- Number of groups:  $I$
- Number of observations in group  $i$ :  $n_i$
- Mean for group  $i$ :  $\bar{Y}_i$
- Standard deviation for group  $i$ :  $s_i$
- Total number of observations:  $n = \sum_{i=1}^I n_i$

- Grand mean (mean of all the observations):

$$\bar{Y} = \frac{\sum_{i=1}^I \sum_{j=1}^{n_i} Y_{ij}}{n} = \frac{\sum_{i=1}^I n_i \bar{Y}_i}{n}$$

- Pooled variance:

$$s^2 = \frac{\sum_{i=1}^I (n_i - 1) s_i^2}{\sum_{i=1}^I (n_i - 1)} = \frac{\sum_{i=1}^I (n_i - 1) s_i^2}{n - I}$$

- Pooled standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^I (n_i - 1) s_i^2}{n - I}}$$

## Example: weight gain of lambs

The following table shows the weight gains (in 2 weeks) of young lambs on three different diets. (These data are fictitious, but are realistic in all respects except for the fact that the group means are whole numbers.)

Table 11.2.1 Weight gains of lambs (lb) *			
	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
$n_i$	3	5	4
Sum = $\sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean = $\bar{y}_i$	11.000	15.000	12.000
SD = $s_i$	4.359	4.950	4.967
*Extra digits are reported for accuracy of subsequent calculations.			

The total number of observations is  $n = \sum_{i=1}^3 n_i = 3 + 5 + 4 = 12$  and the grand mean is

$$\bar{Y} = \frac{\sum_{i=1}^3 n_i \bar{Y}_i}{n} = \frac{3 \times 11 + 5 \times 15 + 4 \times 12}{12} = \frac{156}{12} = 13 \text{ lb.}$$

The pooled variance and standard deviation are calculated as

$$s^2 = \frac{(3-1) \times 4.359^2 + (5-1) \times 4.950^2 + (4-1) \times 4.967^2}{12-3} = \frac{210}{9} = 23.33,$$

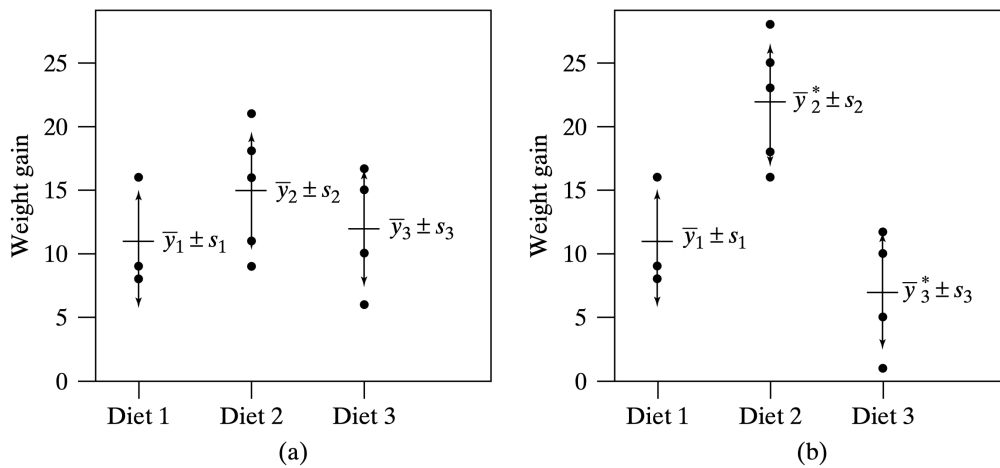
and

$$s = \sqrt{23.33} = 4.83 \text{ lb.}$$

## Variation within groups

- The pooled variance is a weighted average of the group sample variances, and thus a sensible representative value for the variability within groups.
- Note that the pooled variance depends only on the variability within the groups and not on their mean values.
- The pooled variance  $s^2$  is known as the **mean square within groups**, or **MSW**. The numerator of MSW is known as **the sum of squares within groups**, or **SSW**, while the denominator is the **degrees of freedom within groups**.

$$\text{MSW} = \frac{\sum_{i=1}^I (n_i - 1) s_i^2}{n - I} = \frac{\text{SSW}}{n - I}.$$



## Variation between groups

- For two groups, the difference between the groups is simply described by  $(\bar{Y}_1 - \bar{Y}_2)$ . How can we describe between-group variability for more than two groups?
- One naive idea is to simply compute the sample variance of the group means. The **mean square between groups**, or **MSB** is motivated by this idea. Specifically,

$$\text{MSB} = \frac{\sum_{i=1}^I n_i (\bar{Y}_i - \bar{Y})^2}{I - 1} = \frac{\text{SSB}}{I - 1},$$

where the numerator **SSB** is **the sum of squares between groups** and the denominator  $I - 1$  is the **degrees of freedom between groups**.

- The SSB and MSB measure the variability between the sample means of the groups.

For the data in weight gain of lambs example, one has

$$\text{MSW} = s^2 = 23.33, \text{SSW} = 210,$$

and

$$\text{SSB} = 3 \times (11 - 13)^2 + 5 \times (15 - 13)^2 + 4 \times (12 - 13)^2 = 36,$$

$$\text{MSB} = \frac{\text{SSB}}{I - 1} = 18.$$

## A fundamental relationship of ANOVA

The name *analysis of variance* derives from a fundamental relationship involving SSB and SSW. Consider an individual observation  $Y_{ij}$ . It is obviously true that

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

This equation expresses the deviation of an observation from the grand mean as the sum of two parts: a within-group deviation  $(Y_{ij} - \bar{Y}_i)$  and a between-group deviation  $(\bar{Y}_i - \bar{Y})$ . It

is also true (but not at all obvious) that the analogous relationship holds for the corresponding sums of squares; that is

$$\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (\bar{Y}_i - \bar{Y})^2,$$

which, by rewriting each of the sums on the right-hand side, can be expressed as

$$\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^I (n_i - 1)s_i^2 + \sum_{i=1}^I n_i(\bar{Y}_i - \bar{Y})^2 = \text{SSW} + \text{SSB}.$$

The quantity on the left-hand side is called the **total sum of squares**, or **SSTO**:

$$\text{SSTO} = \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2.$$

Note that SSTO measures variability among all  $n$  observations in the  $I$  groups. It follows that

$$\text{SSTO} = \text{SSW} + \text{SSB}.$$

The preceding fundamental relationship shows how *the total variation* in the data set can be analyzed, or broken down, into two interpretable components: *between-sample variation* and *within-sample variation*.

Note that the corresponding degrees of freedom have the same relationship; that is

$$n - 1 = (n - I) + (I - 1),$$

where the left-hand side is called the **total degrees of freedom**.

For the data in weight gain of lambs example, we found  $\bar{Y} = 13$  lb; we calculate SSTO as

$$\begin{aligned} \text{SSTO} &= \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 \\ &= [(8 - 13)^2 + (16 - 13)^2 + (9 - 13)^2] \\ &\quad + [(9 - 13)^2 + (16 - 13)^2 + (21 - 13)^2 + (11 - 13)^2 + (18 - 13)^2] \\ &\quad + [(15 - 13)^2 + (10 - 13)^2 + (17 - 13)^2 + (6 - 13)^2] \\ &= 246. \end{aligned}$$

For these data, we found that  $\text{SSW} = 210$  and  $\text{SSB} = 36$ . We verify that

$$246 = 210 + 36.$$

Also, we found that the degrees of freedom within groups = 9 and the degrees of freedom between groups = 2. We verify that

$$11 = 9 + 2.$$

## The ANOVA Table

When working with the ANOVA quantities, it is customary to arrange them in a table. The table below shows the ANOVA for the lamb weight-gain data. Notice that the ANOVA table clearly shows the additivity of the sums of squares and the degrees of freedom.

**Table 11.2.3** ANOVA table for lamb weight gains

Source	df	SS	MS
Between diets	2	36	18.00
Within diets	9	210	23.33
Total	11	246	

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between groups	$I - 1$	$\sum_{i=1}^I n_i (\bar{Y}_i - \bar{Y})^2$	SS/df
Within groups	$n - I$	$\sum_{i=1}^I (n_i - 1) s_i^2$	SS/df
Total	$n - 1$	$\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$	

## The Analysis of Variance Model

We think of  $Y_{ij}$  as a random observation from group  $i$ , where the population mean of group  $i$  is  $\mu_i$ . It can be helpful to think of ANOVA in terms of the following model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

where

- $\mu$ : grand population mean,
- $\tau_i$ : effect of group  $i$ ,
- $\varepsilon_{ij}$ : random error associated with the  $j$ th observation in group  $i$ .

Thus the preceding model can be stated in words as

$$\text{observation} = \text{overall average} + \text{group effect} + \text{random error}.$$

The group effect  $\tau_i$  can be regarded as the difference between the population mean for group  $i$ ,  $\mu_i$ , and the grand population mean,  $\mu$ . Thus,

$$\tau_i = \mu_i - \mu$$

and the preceding model becomes

$$Y_{ij} = \mu_i + \varepsilon_{ij}.$$

The null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$$

is equivalent to

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_I = 0.$$

If  $H_0$  is false, then at least some of the groups differ from the others. If  $\tau_i$  is positive, then observations from group  $i$  tend to be greater than the overall average; if  $\tau_i$  is negative, then data from group  $i$  tend to be less than the overall average.

The population parameters  $\mu$ ,  $\mu_i$ ,  $\tau_i$ , and  $\varepsilon_{ij}$  can be estimated by the corresponding sample quantities.

$$\hat{\mu} = \bar{Y}, \hat{\mu}_i = \bar{Y}_i, \hat{\tau}_i = \bar{Y}_i - \bar{Y}, \hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_i.$$

Putting these estimates together, we have

$$Y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\varepsilon}_{ij} = \bar{Y} + (\bar{Y}_i - \bar{Y}) + (Y_{ij} - \bar{Y}_i).$$

While the terms "between-groups" and "within-groups" are not technical terms, they are useful in describing and understanding the ANOVA model. Computer software and other texts commonly refer to these sources of variability as **treatment (between groups)** and **error (within groups)**.

For the data in weight gain of lamb example, the estimate of the grand population mean is  $\hat{\mu} = 13$ . The estimated group effects are

$$\hat{\tau}_i = \bar{Y}_i - \bar{Y} = 11 - 13 = -2, \hat{\tau}_2 = 15 - 13 = 2, \hat{\tau}_3 = 12 - 13 = -1.$$

Thus, we estimate that Diet 2 increases weight gain by 2 lb on average (when compared to the average of the three diets), Diet 1 decreases weight gain by an average of 2 lb, and Diet 3 decreases weight gain by 1 lb, on average.

When we conduct an analysis of variance, we are comparing the sizes of the sample group effects, the  $\hat{\tau}_i$ 's, to the sizes of the random errors in the data, the  $\hat{\varepsilon}_{ij}$ 's. We can see that

$$SSB = \sum_{i=1}^I n_i \hat{\tau}_i^2, \quad SSW = \sum_{i=1}^I \sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^2.$$

## The Global $F$ Test

The global null hypothesis is

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$$

against the alternative hypothesis

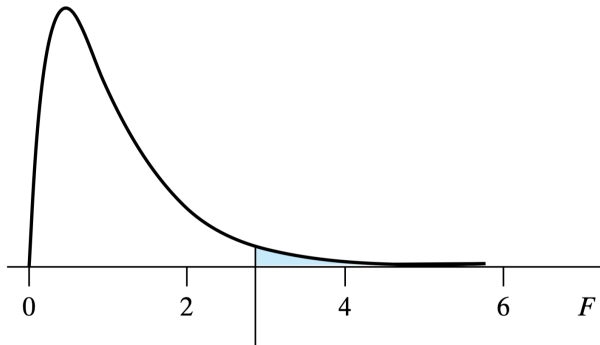
$$H_A : \text{The } \mu_i \text{'s are not all equal.}$$



Note that  $H_0$  is compound (unless  $I = 2$ ), and so rejection of  $H_0$  does not specify which  $\mu_i$ 's are different. If we reject  $H_0$ , then we conduct a further analysis to make detailed comparisons among the  $\mu_i$ 's.

## The $F$ distribution

The form of an  $F$  distribution depends on two parameters: the **numerator degrees of freedom** and the **denominator degrees of freedom**. Critical values for the  $F$  distribution are given in  $F$  Table. Note that  $F$  Table occupies 10 pages, each page having a different value of the numerator df. As a specific example, for numerator df = 4 and denominator df = 20, we find in  $F$  Table that  $F_{4,20}(0.05) = 2.87$ ; this value is shown in the figure below.



The  $F$  test is a classical test of the preceding global null hypothesis. The test statistic, the  $F$  statistic, is calculated as follows:

$$T = \frac{\text{MSB}}{\text{MSW}}.$$

From the definitions of the mean squares, it is clear that  $T$  will be large if the discrepancies among the group means ( $\bar{Y}_i$ 's) are large relative to the variability within the groups. Thus, large values of  $T$  tend to provide evidence against  $H_0$  (evidence for a difference among the group means).

It can be shown mathematically that the null distribution of the test statistic  $T$  is the  $F$  distribution with the numerator df being the df between groups and the denominator df being the df within groups. Specifically,

$$T \stackrel{H_0}{\sim} F_{I-1, n-I}.$$

Therefore,  $H_0$  is rejected at the  $\alpha$  level of significance if

$$p\text{-value} = P(F_{I-1, n-I} > T) < \alpha \text{ or } T > F_{I-1, n-I}(\alpha).$$

For the data in weight gain of lamb example, the global null hypothesis and alternative can be stated verbally as

$H_0$ : Mean weight gain is the same on all three diets. v.s.  $H_A$ : Mean weight gain is not the same on all three diets.

or symbolically as

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ v.s. } H_A : \text{ The } \mu_i \text{'s are not all equal.}$$

From the ANOVA table we find

$$T = \frac{18}{23.33} = 0.77.$$

The degrees of freedom can also be read from the ANOVA table as numerator  $df = 2$  and denominator  $df = 9$ . From  $F$  Table we find  $F_{2,9}(0.20) = 1.93$ . So  $p\text{-value} > 0.20$  (Computer software gives  $p\text{-value} = 0.4907$ ). Thus, there is a lack of significant evidence against  $H_0$ ; there is insufficient evidence to conclude that there is any difference among the diets with respect to population mean weight gain.

## Linear Combinations of Means

In many studies, interesting questions can be addressed by considering linear combinations of the group means. A linear combination  $L$  is a quantity of the form

$$L = \sum_{i=1}^I m_i \bar{Y}_i,$$

where the  $m_i$ 's are the multipliers of the  $\bar{Y}_i$ 's.

### Standard error of a linear combination

Each linear combination  $L$  is an estimate, based on the  $\bar{Y}_i$ 's, of the corresponding linear combination of the population means ( $\mu_i$ 's). As a basis for statistical inference, we need to consider the standard error of a linear combination, which is calculated as follows.

The standard error of the linear combination

$$L = \sum_{i=1}^I m_i \bar{Y}_i$$

is

$$SE_L = \sqrt{\text{MSW} \times \sum_{i=1}^I \frac{m_i^2}{n_i}}.$$

### Confidence intervals

Linear combinations of means can be used for testing hypotheses and for constructing confidence intervals. Critical values are obtained from Student's  $t$  distribution with  $df$  being

the degrees of freedom within group, i.e.,  $n - I$ . Confidence intervals are constructed using the familiar Student's  $t$  format.

In general, a  $1 - \alpha$  confidence interval for the linear combination  $L$  is

$$L \pm t_{n-I}(\alpha/2) \times SE_L.$$

## Multiple Comparisons

After finding significant evidence for a difference among population means  $\mu_1, \mu_2, \dots, \mu_I$  using a global  $F$  test, we wish to conduct pairwise comparisons between different population means to further detect where the difference lies in. However, repeated  $t$  tests can lead to an increased overall risk of Type I error. The **Bonferroni's method** is one popular method to control the overall risk of Type I error.

The Bonferroni's method is based on a very simple and general relationship: *The probability that at least one of several events will occur cannot exceed the sum of the individual probabilities.* For instance, suppose we conduct five tests of hypotheses, each at  $\alpha_i = 0.01$ . Then the overall risk of Type I error  $\alpha$  (the chance of rejecting at least one of the six hypotheses when in fact all of them are true) cannot exceed  $5 \times 0.01 = 0.05$ .

Turning this logic around, suppose an investigator plans to conduct five tests of hypotheses and wants the overall risk of Type I error not to exceed  $\alpha = 0.05$ . A conservative approach is to conduct each of the separate tests at the significance level  $\alpha_i = 0.05/5 = 0.01$ ; this is called a **Bonferroni adjustment**.

A Bonferroni adjustment can also be made for confidence intervals. For instance, suppose we wish to construct five confidence intervals and desire an overall probability of 95% that all the intervals contain their respective parameters ( $\alpha = 0.05$ ). Then this can be accomplished by constructing each interval at confidence level 99% (because  $0.05/5 = 0.01$  and  $1 - 0.01 = 0.99$ ).

In general, to construct  $k$  Bonferroni-adjusted confidence intervals with an overall probability of  $100(1 - \alpha)\%$  that all the intervals contain their respective parameters, we construct each interval at confidence level  $100(1 - \alpha/k)\%$ . Formally, the Bonferroni-adjusted  $1 - \alpha$  confidence interval for  $\mu_a - \mu_b$  is

$$(\bar{Y}_a - \bar{Y}_b) \pm t_{n-I}(\alpha/(2k)) \times SE_{\bar{Y}_a - \bar{Y}_b},$$

where the standard error

$$SE_{\bar{Y}_a - \bar{Y}_b} = \sqrt{MSW \times \left( \frac{1}{n_a} + \frac{1}{n_b} \right)}.$$

Note that the application of Bonferroni's method requires unusual critical values, so standard tables are not sufficient. Bonferroni Table provides Bonferroni multipliers for

confidence intervals that are based on a  $t$  distribution.

## Example: oysters and seagrass

In a study to investigate the effect of oyster density on seagrass biomass, researchers introduced oysters to thirty 1-m<sup>2</sup> plots of healthy seagrass. At the beginning of the study the seagrass was clipped short in all plots. Next, 10 randomly chosen plots received a high density of oysters; 10, an intermediate density; and 10, a low density. As a control, an additional 10 randomly chosen clipped 1-m<sup>2</sup> plots received no oysters. After 2 weeks, the belowground seagrass biomass was measured in each plot (g/m<sup>2</sup>). Data from some plots are missing. A summary of the data as well as the ANOVA table follow.

Table 11.9.1 Belowground seagrass biomass (g/m <sup>2</sup> )				
	Oyster density			
	None (1)	Low (2)	Intermediate (3)	High (4)
Mean	34.81	33.13	28.33	15.00
SD	13.44	17.36	17.11	10.97
$n$	9	10	8	10

Table 11.9.2 ANOVA summary of belowground seagrass biomass (g/m <sup>2</sup> )					
	df	Sum of squares	Mean squares	$F$	$P$ -value
Between	3	2365.5	788.51	3.5688	0.0243
Within	33	7291.1	220.94		
Total	36	9656.6			

The  $p$ -value for the global  $F$  test is 0.0243, indicating that there is significant evidence of a difference among the biomass means under these experimental conditions. We thus proceed with pairwise comparisons to further detect which two conditions are different. To control the overall risk of Type I error, we calculate the Bonferroni-adjusted 95% confidence intervals for the total of six comparisons. Each individual confidence interval shall be constructed at confidence level 99.17% since  $0.05/6 = 0.0083$  and  $1 - 0.0083 = 0.9917$ .

The following table summarizes the Bonferroni-adjusted confidence intervals for the total six pairwise comparisons.

Table 11.9.5 Intermediate computations and experimentwise 95% (99.17% comparisonwise) Bonferroni intervals comparing belowground biomass under different oyster density conditions					
Comparison	$d_{ab} = \bar{y}_a - \bar{y}_b$	$SE_{D_{ab}}$	$t_{30,0.025/6} \times SE_{D_{ab}}$	Lower 99.17%	Upper 99.17%
None–low	1.68	6.828	19.289	–17.6	21.0
None–intermediate	6.48	7.221	20.399	–13.9	26.9
<i>None–high</i>	<i>19.81</i>	6.828	<i>18.775</i>	<i>0.5</i>	<i>39.1</i>
Low–intermediate	4.80	7.049	19.913	–15.1	24.7
Low–high	18.13	6.646	18.775	–0.6	36.9
Intermediate–high	13.33	7.049	19.913	–6.6	33.2

\*Intervals not containing zero (i.e., where there is a statistically significant difference between the group means) are in italics. Note the first two columns ( $d_{ab}$  and  $SE_{D_{ab}}$ ) are identical to those presented in Table 11.9.3.

- Unfortunately, the Bonferroni intervals are often **overly conservative** so that the actual value of  $\alpha$  is much less than the desired overall risk of Type I error, and thus too much power is sacrificed for Type I error protection. More complex procedures such as **Fisher's Least Significant Difference** and **Tukey's Honest Significant Difference** are able to achieve higher power than Bonferroni.
- An advantage of the Bonferroni method is that it is widely applicable and can easily be generalized to situations beyond ANOVA.

## Conditions of ANOVA

The ANOVA techniques described in this chapter, including the global F test, are valid if the following conditions hold.

- Design conditions
  - It must be reasonable to regard the groups of observations as **random samples** from their respective populations.
  - The  $I$  samples must be **independent** of each other.
- Population conditions
  - The  $I$  population distributions must be (approximately) **normal** with **equal standard deviations**:

$$\sigma_1 = \sigma_2 = \cdots = \sigma_I.$$