Chapter 9

Analysis of variance

Chapter outline

Objective

The objective of this chapter is to analyze the means of several populations by identifying the sources of variability of the data.

John Wilder Tukey
(Source: http://en.wikipedia.org/wiki/John_Tukey).

John W. Tukey (1915–2000), a chemist turned topologist turned statistician, was one of the most influential statisticians of the past 50 years. He is credited with inventing the word *software*. He worked as a professor at Princeton University and a senior researcher at AT&T's Bell Laboratories. He made significant contributions to the fields of exploratory data analysis and robust estimation. His works on the spectrum analysis of time series and other aspects of

Mathematical Statistics with Applications in R. https://doi.org/10.1016/B978-0-12-817815-7.00009-9

digital signal processing have been widely used in engineering and science. He coined the word *bit*, which refers to a unit of information processed by a computer. In collaboration with Cooley, in 1965, Tukey introduced the fast Fourier transform (FFT) algorithm that greatly simplified computation for Fourier series and integrals. Tukey authored or coauthored many books on statistics and wrote more than 500 technical papers. Among Tukey's most far-reaching contributions was his development of techniques for "robust analysis," an approach to statistics that guards against wrong answers in situations where a randomly chosen sample of data happens to poorly represent the rest of the data set. Tukey also made significant contributions to the analysis of variance.

9.1 Introduction

Suppose that we are interested in the effects of four different types of chemical fertilizers on the yield of rice, measured in pounds per acre. If there is no difference between the different types of fertilizers, then we would expect all the mean yields to be approximately equal. Otherwise, we would expect the mean yields to differ. The different types of fertilizers are called treatments and their effects are the treatment effects. The yield is called the response. Typically, we have a model with a response variable that is possibly affected by one or more treatments. The study of these types of models falls under the purview of design of experiments, which we discussed in Chapter 8. In this chapter we concentrate on the analysis aspect of the data obtained from the designed experiments. If the data came from one or two populations, we could use the techniques learned in Chapters 5 and 6. Here, we introduce some tests that are used to analyze the data from more than two populations. These tests are used to deal with treatment effects, including tests that take into account other factors that may affect the response. The hypothesis that the population means are equal is considered equivalent to the hypothesis that there is no difference in treatment effects. The analytical method we will use in such problems is called the analysis of variance (ANOVA). The initial development of this method could be credited to Sir Ronald A. Fisher, who introduced this method for the analysis of agricultural field experiments. The "green revolution" in agriculture would have been impossible without the development of the theory of experimental design and the methods of ANOVA.

ANOVA is one of the most flexible and practical techniques for comparing several means. It is important to observe that ANOVA is not about analyzing the population variance. In fact, we are analyzing treatment means by identifying sources of variability of the data. In its simplest form, ANOVA can be considered as an extension of the test of hypothesis for the equality of two means that we learned in Chapter 6. Actually, the so-called one-way ANOVA is a generalization of the two-means procedure to a test of equality of the means of more than two independent, normally distributed populations.

Recall that the methods of testing H_0: $\mu_1 - \mu_2 = 0$, such as the t-test, were discussed earlier. In this chapter, we are concerned with studying situations involving the comparison of more than two population or treatment means. For example, we may be interested in the question, Do the rates of heart attack and stroke differ for three different groups of people with high cholesterol levels (borderline high, such as 150−199 mg/dL; high, such as 200−239 mg/dL; very high, such as greater than 240 mg/dL) and a control group given different dosage levels of a particular cholesterol-lowering drug (say, a particular statin drug)? Let us consider four populations with means μ_1, μ_2, μ_3, and μ_4, and say that we wish to test the hypothesis $\mu_1 = \mu_2 = \mu_3 = \mu_4$. That is, the true mean rate is the same for all four groups. The question here is, Why do we need a new method to test for differences among the four procedure population means? Why not use z- or t-tests for all possible pairs and test for differences in each pair? If any one of these tests leads to the rejection of the hypothesis of equal means, then we might conclude that at least two of the four population means differ. The problem with this approach is that our final decision is based on results of $\binom{4}{2} = 6$ different tests, and any one of them can be wrong. For each of the six tests, let $\alpha = 0.10$ be the probability of being wrong (type I error). Then the probability that at least one of the six tests leads to the conclusion that there is a difference leads to an error of $1 - (0.9)^6 = 0.46856$, which clearly is much larger than 0.10, thus resulting in a large increase in the type I error rate. Hence, if an ordinary t-test is used to make several treatment comparisons from the same data, the actual α-value applying to the tests taken as a group will be larger than the specified value of α, and one is likely to declare significance when there is none.

ANOVA procedures were developed to eliminate the increase in error rates resulting from multiple t-tests. With ANOVA, we are able to set one α level and test whether any of the group means differ from one another. Given a sample from each of the populations, our interest is to answer the question, Are the observed discrepancies among the different sample means merely due to chance fluctuations, or are they due to inherent differences among the populations? ANOVA separates the effect of purely random variations from those caused by existing differences among population means. The phrase "analysis of variance" springs from the idea of analyzing variability in the data to see how much can be attributed to differences in μ and how much is due to variability in the individual populations. The ANOVA method incorporates information on variability from all of the samples simultaneously. At the heart of ANOVA is the fact that variances can be partitioned, with each partition attributable to a specific source. The method inspects various sums of squares (which are measures of variation in a sample) calculated from the data. ANOVA looks at two types of sums of squares: sums of squares within groups and sums of squares between groups. That is, it looks at each of the distributions and compares the between-group differences (variation in group means) with the within-group differences (variation in individuals' scores within groups).

9.2 Analysis of variance method for two treatments (optional)

In this section, we present the simplest form of the ANOVA procedure, the case of studying the means of two populations, I and II. For comparing only two means, the ANOVA will result in the same conclusions as the t-test for independent random samples. The basic purpose of this section is to introduce the concept of ANOVA in simpler terms. Let us consider two random samples of size n_1 and n_2, respectively. That is, $y_{11}, y_{12}, \ldots, y_{1n_1}$ from population I and $y_{21}, y_{22}, \ldots, y_{2n_2}$ from population II. Let

$$\overline{y}_1 = \frac{y_{11} + y_{12} + \cdots + y_{1n_1}}{n_1} \text{ (sample mean from population I)},$$

and

$$\overline{y}_2 = \frac{y_{21} + y_{22} + \cdots + y_{2n_2}}{n_2} \text{ (sample mean from population II)}.$$

These samples are assumed to be independent and come from normal populations with respective means μ_1, μ_2, and variances $\sigma_1^2 = \sigma_2^2$. We wish to test the hypothesis:

$$H_0: \mu_1 = \mu_2 \text{ vs. } H_a: \mu_1 \neq \mu_2.$$

The total variation of the two combined response measurements about $\overline{y}$ (the sample mean of all $n = n_1 + n_2$ observations) is (SS is used for sum of squares) defined as:

$$Total\ SS = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(y_{ij} - \overline{y}\right)^2. \tag{9.1}$$

That is,

$$\overline{y} = \frac{y_{11} + y_{12} + \cdots + y_{1n_1} + y_{21} + y_{22} + \cdots + y_{2n_2}}{n} = \frac{1}{n}\sum_{ij} y_{ij}.$$

The total sums of squares measure the total spread of scores around the grand mean, $\overline{y}$. We can rewrite Eq. (9.1) as:

$$\begin{aligned}
Total\ SS &= \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left(y_{ij} - \overline{y}\right)^2 \\
&= \sum_{j=1}^{n_1} \left(y_{1j} - \overline{y}\right)^2 + \sum_{j=1}^{n_2} \left(y_{2j} - \overline{y}\right)^2 \\
&= \sum_{j=1}^{n_1} \left(y_{1j} - \overline{y}_1 + \overline{y}_1 - \overline{y}\right)^2 + \sum_{j=1}^{n_2} \left(y_{2j} - \overline{y}_2 + \overline{y}_2 - \overline{y}\right)^2 \\
&= \sum_{j=1}^{n_1} \left(y_{1j} - \overline{y}_1\right)^2 + n_1\left(\overline{y}_1 - \overline{y}\right)^2 + 2\left(\overline{y}_1 - \overline{y}\right) \sum_{j=1}^{n_1} \left(y_{1j} - \overline{y}_1\right) \\
&+ \sum_{j=1}^{n_2} \left(y_{2j} - \overline{y}_2\right)^2 + n_2\left(\overline{y}_2 - \overline{y}\right)^2 + 2\left(\overline{y}_2 - \overline{y}\right) \sum_{j=1}^{n_2} \left(y_{2j} - \overline{y}_2\right).
\end{aligned}$$

Note that $\sum_{j=1}^{n_1}\left(y_{1j} - \overline{y}_1\right) = 0 = \sum_{j=1}^{n_2}\left(y_{2j} - \overline{y}_2\right)$. Thus, we obtain:

$$
\begin{aligned}
Total\ SS &= \sum_{j=1}^{n_1}\left(y_{1j} - \overline{y}_1\right)^2 + \sum_{j=1}^{n_2}\left(y_{2j} - \overline{y}_2\right)^2 \\
&\quad + n_1\left(\overline{y}_1 - \overline{y}\right)^2 + n_2\left(\overline{y}_2 - \overline{y}\right)^2 \qquad (9.2) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}_i\right)^2 + \sum_{i=1}^{2} n_i\left(\overline{y}_i - \overline{y}\right)^2.
\end{aligned}
$$

Define SST, the sum of squares for a treatment, as:

$$SST = \sum_{i=1}^{2} n_i\left(\overline{y_i} - \overline{y}\right)^2.$$

The *SST* measures the total spread of the group means $\overline{y}_i$ with respect to the grand mean, $\overline{y}$. Also, *SSE* represents the *sum of squares of errors* given by:

$$
\begin{aligned}
SSE &= \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}_i\right)^2 \\
&= \sum_{j=1}^{n_1}\left(y_{1j} - \overline{y}_1\right)^2 + \sum_{j=1}^{n_2}\left(y_{2j} - \overline{y}_2\right)^2 \\
&= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2,
\end{aligned}
$$

where s_1^2 and s_2^2 are the unbiased sample variances of the two random samples. Note that this connects the sum of squares to the concept of variance we have been using in previous chapters. We can now rewrite Eq. (9.2) as:

$$Total\ SS = SSE + SST.$$

It should be clear that the *SSE* measures the within-sample variation of the y values (effects), whereas *SST* measures the variation among the two sample means. The logic by which the ANOVA tests is as follows: If the null hypothesis is true, then SST compared with *SSE* should be about the same, or less. The larger the *SST*, the greater will be the weight of evidence to indicate a difference in the means μ_1 and μ_2. The question then is, how large?

To answer this question, let us suppose we have two populations that are normal. That is, let Y_{ij} be $N(\mu_i, \sigma^2)$ distributed with values y_{ij}. Then, the pooled unbiased estimate of σ^2 is given by:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{SSE}{n_1 + n_2 - 2}.$$

Hence,

$$\sigma^2 = E\left(S_p^2\right) = E\left(\frac{SSE}{n_1 + n_2 - 2}\right).$$

Also, we can write:

$$\frac{SSE}{\sigma^2} = \sum_{j=1}^{n_1}\frac{\left(Y_{1j} - \overline{Y_1}\right)^2}{\sigma^2} + \sum_{j=1}^{n_2}\frac{\left(Y_{2j} - \overline{Y_2}\right)^2}{\sigma^2},$$

which has a χ^2 distribution with $(n_1 + n_2 - 2)$ degrees of freedom.
Under the hypothesis that $\mu_1 = \mu_2$, $E(\mathrm{SST}) = \sigma^2$. Furthermore,

$$Z = \frac{\overline{Y_1} - \overline{Y_2}}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1).$$

This implies that:

$$Z^2 = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left[\frac{\left(\overline{Y_1}-\overline{Y_2}\right)^2}{\sigma^2}\right] = \frac{SST}{\sigma^2},$$

has a χ^2 distribution with 1 degree of freedom. It can be shown that *SST* and *SSE* are independent. From Chapter 4, we restate the following result.

Theorem 9.2.1. *If χ_1^2 has υ_1 degrees of freedom, χ_2^2 has υ_2 degrees of freedom, and χ_1^2 and χ_2^2 are independent, then* $\mathrm{F} = \frac{\chi_1^2/\upsilon_1}{\chi_2^2/\upsilon_2}$ *has an* F*-distribution with υ_1 numerator degrees of freedom and υ_2 denominator degrees of freedom.*

Using the foregoing result, we have:

$$\frac{SST/(1)\sigma^2}{SSE/(n_1+n_2-2)\sigma^2} = \frac{SST/1}{SSE/(n_1+n_2-2)},$$

which has an *F*-distribution with $\upsilon_1 = 1$ numerator degrees of freedom and $\upsilon_2 = (n_1 + n_2 - 2)$ denominator degrees of freedom.

Now, we introduce the mean square error (MSE), defined as:

$$MSE = \frac{SSE}{(n_1+n_2-2)}$$

$$= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)},$$

and the mean square treatment (MST), given by:

$$MST = \frac{SST}{1}$$

$$= \left[n_1\left(\overline{y}_1-\overline{y}\right)^2 + n_2\left(\overline{y}_2-\overline{y}\right)^2\right].$$

Under the null hypothesis, H_0: $\mu_1 = \mu_2$, both MST and MSE estimate σ^2 without bias. When H_0 is false and $\mu_1 \neq \mu_2$, MST estimates something larger than σ^2 and will be larger than MSE. That is, if H_0 is false, then $E(MST) > E(MSE)$ and the greater the differences among the values of μ, the larger $E(MST)$ will be relative to $E(MSE)$.

Hence, to test $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$, we use the *F*-test, given by:

$$F = \frac{MST}{MSE},$$

as the test statistic. Thus, for a given α, the rejection region is $\{F > F_\alpha\}$. It is important to observe that compared with the small sample *t*-test, here we work with variability. Now we summarize the ANOVA procedure for the two-sample case.

Analysis of variance procedure for two treatments

For equal sample sizes $n = n_1 = n_2$, assume $\sigma_1^2 = \sigma_2^2$.

We test:

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2.$$

1. Calculate $\overline{y}_1, \overline{y}_2, \sum_{ij} y_{ij}^2, \sum_{ij} y_{ij}$, and find:

$$SST = \sum_{i=1}^{2} n_i\left(\overline{y}_i - \overline{y}\right)^2.$$

Also calculate:

$$Total\ SS = \sum_i\sum_j y_{ij}^2 - \frac{\left(\sum_i\sum_j y_{ij}\right)^2}{n_1+n_2}.$$

Then:

$$SSE = Total\ SS - SST.$$

Continued

Analysis of variance procedure for two treatments—cont'd

2. Compute:

$$MST = \frac{SST}{1},$$

and

$$MSE = \frac{SSE}{n_1 + n_2 - 2}.$$

3. Compute the test statistic,

$$F = \frac{MST}{MSE}.$$

4. For a given α, find the rejection region as:

$$RR: F > F_\alpha,$$

based on 1 numerator and $(n_1 + n_2 - 2)$ denominator degrees of freedom.

5. **Conclusion:** If the test statistic F falls in the rejection region, conclude that the sample evidence supports the alternative hypothesis that the means are indeed different for the two treatments.

Assumptions: The populations are normal with equal but unknown variances.

EXAMPLE 9.2.1

The following data represent a random sample of end-of-year bonuses for lower-level managerial personnel employed by a large firm. Bonuses are expressed in percentage of yearly salary.

Female	6.2	9.2	8.0	7.7	8.4	9.1	7.4	6.7
Male	8.9	10.0	9.4	8.8	12.0	9.9	11.7	9.8

The objective is to determine whether the male and female bonuses are the same. We can answer this question by connecting the following:

(a) Use the ANOVA approach to test the appropriate hypothesis. Use $\alpha = 0.05$.
(b) What assumptions are necessary for the test in (a)?
(c) Test the appropriate hypothesis by using the two-sample t-test for comparing population means. Compare the value of the t-statistic with the value of the F-statistic calculated in (a).

Solution

(a) *We need to test:*

$$H_0: \mu_1 = \mu_2 \text{ vs. } H_a: \mu_1 \neq \mu_2.$$

From the random samples, we obtain the following needed estimates, $n_1 = n_2 = 8$:

$$\overline{y}_1 = 7.8375,\ \overline{y}_2 = 10.0625, \sum_{ij} y_{ij}^2 = 1319.34, \sum_{ij} y_{ij} = 143.20,\ \overline{y} = 8.95$$

and

$$SST = \sum_{i=1}^{2} n_i \left(\overline{y}_i - \overline{y}\right)^2 = 19.8025.$$

Therefore,

$$\text{Total } SS = \sum_i \sum_j y_{ij}^2 - \frac{\left(\sum_i \sum_j y_{ij}\right)^2}{(n_1 + n_2)}$$

$$= 1391.34 - \frac{(143.2)^2}{16}$$

$$= 109.70.$$

Then:

$$SSE = Total\ SS - SST$$
$$= 109.7 - 19.8025 = 89.8975,$$
$$MST = \frac{SST}{1} = 19.8025,$$

and

$$MSE = \frac{SSE}{2n_1 - 2}$$
$$= \frac{89.8975}{14}$$
$$= 6.42125.$$

Hence, the test statistic:

$$F = \frac{MST}{MSE}$$
$$= \frac{19.8025}{6.42125}$$
$$= 3.0839.$$

For $\alpha = 0.05$, $F_{0.05,1,14} = 4.60$. *Hence, the rejection region is* $\{F > 4.60\}$. *Because* 3.0839 *is not greater than* 4.60, H_0 *is not rejected. There is not enough evidence to indicate that the average percentage bonuses are different for men and women at* $\alpha = 0.05$.

(b) *To solve the problem, we assumed that the samples are random and independent with* $n_1 = n_2 = 8$, *drawn from two normal populations with means* μ_1 *and* μ_2 *and common variance* σ^2.

(c) *The value of MSE is the same as* $s^2 = s_p^2 = 6.42125$. *Also,* $\bar{y}_1 = 7.8375$ *and* $\bar{y}_2 = 10.0625$. *Then, the* t-*statistic is:*

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{7.8375 - 10.0625}{\sqrt{6.42125\left(\frac{1}{8} + \frac{1}{8}\right)}} = -1.756.$$

Now, $t_{0.025,\ 14} = 2.145$ *and hence, the rejection region is* $\{t < -2.145\}$.

Because -1.756 *is not less than* -2.145, H_0 *is not rejected, which implies that there is no significant difference between the bonuses for males and females. Note also that* $t^2 = F$, *that is,* $(-1.756)^2 = 3.083$, *implying that in the two-sample case, the* t-*test and* F-*test lead to the same result.*

It is not surprising that, in the previous example, the conclusions reached using ANOVA and two-sample *t*-tests are the same. In fact, it can be shown that for two sets of independent and normally distributed random variables, the two procedures are entirely equivalent for a two-sided hypothesis. However, a *t*-test can also be applied to a one-sided hypothesis, whereas ANOVA cannot. The purpose of this section is only to illustrate the computations involved in the ANOVA procedures as opposed to simple *t*-tests. The ANOVA procedure is effectively used for three or more populations, which is described in the next section.

Exercises 9.2

9.2.1. The following information was obtained from two independent samples selected from two normally distributed populations with unknown but equal standard deviations. Do the data present sufficient evidence to indicate that there is a difference in the mean for the two populations?

Sample 1	1	2	3	3	1	2	1	3	1
Sample 2	2	5	2	4	3	1	2	3	3

(a) Use the ANOVA approach to test the appropriate hypotheses. Use $\alpha = 0.05$.
(b) Test the appropriate hypothesis by using the two-sample t-test for comparing population means. Compare the value of the t-statistic to the value of the F-statistic calculated in (a).

9.2.2. The following information was obtained from two independent samples selected from two normally distributed populations with unknown but equal standard deviations. Do the data present sufficient evidence to indicate that there is a difference in the mean for the two populations?

Sample 1	15	13	11	14	10	12	7	12	11	14	15	
Sample 2	18	16	13	21	16	19	15	18	19	20	21	14

(a) Use the ANOVA approach to test the appropriate hypotheses. Use $\alpha = 0.01$.
(b) Test the appropriate hypothesis by using the two-sample t-test for comparing population means. Compare the value of the t-statistic to the value of the F-statistic calculated in (a).

9.2.3. A company claims that its medicine, brand A, provides faster relief from pain than another company's medicine, brand B. A random sample from each brand gave the following times (in minutes) for relief. Do the data present sufficient evidence to indicate that there is a difference in the mean time to relief for the two populations?

Brand A	47	51	45	53	41	55	50	46	45	51	53	50	48
Brand B	44	48	42	45	44	42	49	46	45	48	39	49	

(a) Use the ANOVA approach to test the appropriate hypothesis. Use $\alpha = 0.01$.
(b) What assumptions are necessary for the conclusion in (a)?
(c) Test the appropriate hypothesis by using the two-sample t-test for comparing population means. Compare the value of the t-statistic to the value of the F-statistic calculated in (a).

9.2.4. Table 9.1 gives mean SAT scores for math by state from 1989 and 1999 for 20 randomly selected states. (Source: *The World Almanac and Book of Facts, 2000*.)

TABLE 9.1 Mean SAT Scores for Math by State.

State	1989	1999
Arizona	523	525
Connecticut	498	509
Alabama	539	555
Indiana	487	498
Kansas	561	576
Oregon	509	525
Nebraska	560	571
New York	496	502
Virginia	507	499
Washington	515	526
Illinois	539	585
North Carolina	469	493
Georgia	475	482
Nevada	512	517
Ohio	520	568
New Hampshire	510	518

Using the ANOVA procedure, test if the mean SAT score for math in 1999 is greater than that in 1989 at $\alpha = 0.05$. Assume that the variances are equal and the samples come from a normal distribution.

9.2.5. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two sets of independent, normally distributed random variables with means μ_1 and μ_2 and the common variance σ^2. Show that the two-sample t-test and the ANOVA are equivalent for testing H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$.

9.3 Analysis of variance for a completely randomized design

In this section, we study the hypothesis-testing problem of comparing population means for more than two independent populations, where the data are about several independent groups (different treatments being applied or different populations being sampled). We have seen in Chapter 8 that the random selection of independent samples from k populations is known as a completely randomized experimental design or one-way classification.

Let $\mu_1, \ldots, \mu_k$ be the means of k normal populations with unknown but equal variance σ^2. The question is whether the means of these groups are different or are all equal. The idea is to consider the overall variability in the data. We partition the variability into two parts: (1) between-groups variability and (2) within-group variability. If between groups is much larger than that within groups, this will indicate that differences between the groups are real, not merely due to the random nature of sampling. Let independent samples be drawn of sizes n_i, $i = 1, 2, \ldots, k$, and let $N = n_1 + \ldots + n_k$. Let y_{ij} be the measured response on the jth experimental unit in the ith sample. That is, Y_{ij} is the jth observation from population i, $i = 1, 2, \ldots, k$, and $j = 1, 2, \ldots, n_i$. Let $\overline{y}$ be the overall mean of all observations. The problem can be formulated as a hypothesis-testing problem, where we need to test:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \text{ vs. } H_a: \text{Not all the } \mu_i's \text{ are equal.}$$

The method of ANOVA tests the null hypothesis H_0 by comparing two unbiased estimates of the variance, σ^2, an estimate based on variations from sample to sample, and the other one based on variations within the samples. We will be rejecting H_0 if the first estimate is significantly larger than the second, so that the samples cannot be assumed to come from the same population. That is, the variances influence the decision.

We can write the total sum of squares of deviations of the response measurements about their overall mean for the k samples into two parts, from the treatment (SST) and from the error (SSE). This partition gives the fundamental relationship in ANOVA, where total variation is divided into two portions: between-sample variation and within-sample variation. That is,

$$Total\ SS = SST + SSE.$$

The following derivations will make computation of these quantities simpler. The total SS can be written as:

$$Total\ SS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - 2\overline{y}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} + N\overline{y}^2.$$

Note that $\overline{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{N}$, and then we have:

$$Total\ SS = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - CM,$$

where CM is the correction factor for the correction for the means and is given by:

$$CM = \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}\right)^2}{N} = N\overline{y}^2.$$

Let

$$T_i = \sum_{j=1}^{n_i} y_{ij}, \text{ be the sum of all the observations in the } i\text{th sample}$$

and

$$\overline{T_i} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, \text{the mean of the observations in the } i\text{th sample.}$$

We can rewrite $\overline{y}$ as:

$$\overline{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{N} = \frac{\sum_{i=1}^{k} n_i \overline{T_i}}{N}.$$

Now, we introduce *SST*, the sum of squares for treatment (sometimes known as between-group sum of squares, SSB) as:

$$SST = \sum_{i=1}^{k} n_i \left(\overline{T_i} - \overline{y}\right)^2.$$

We note that $\left(\overline{T_i}\right)$ is the mean response due to its *i*th treatment and $\overline{y}$ is the overall mean. A large value of $\left(\overline{T_i} - \overline{y}\right)$ is likely to be caused by the *i*th treatment effect being very different from the rest. Hence, *SST* can be used to measure the differences in the treatment effects.

Thus, the *SSE* is given by:

$$SSE = Total\ SS - SST.$$

We must state that the *SSE* is the sum of squares within groups (thus, sometimes *SSE* is referred to as the *within-group sum of squares*, SSW) and this can be seen from rewriting the expression as:

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \overline{T_i}\right)^2.$$

The decomposition of the total sum of squares can be easily seen in Fig. 9.1.

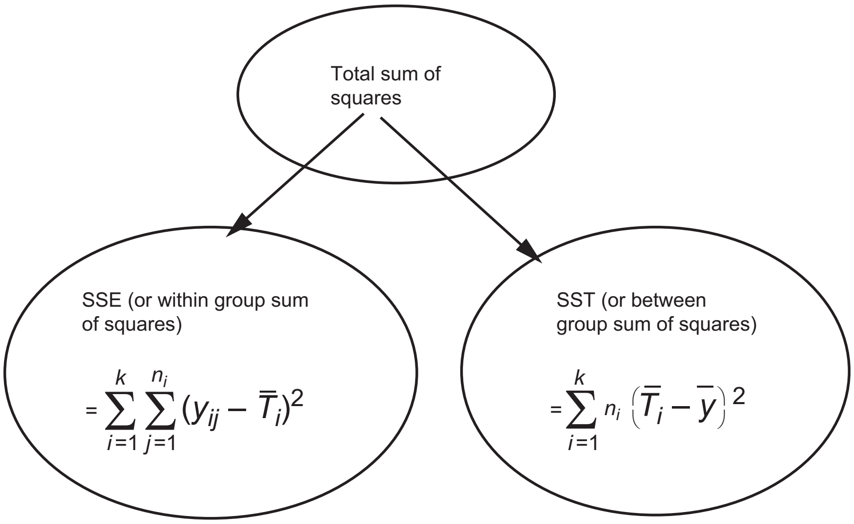


FIGURE 9.1 Decomposition of total sum of squares.

Fig. 9.2 represents one point for each observation against each sample, with SM representing the sample means and GM representing the grand mean. The dotted lines between the SMs and the GM are the distance between them. Taking these distances, squaring, multiplying by the corresponding sample sizes, and summing, we get *SST*. To obtain *SSE*, we take the distances from each group mean, SM, to each member of the group, square them, and add them. In addition, to give an idea of within-group variations, it is customary to draw side-by-side box plots.

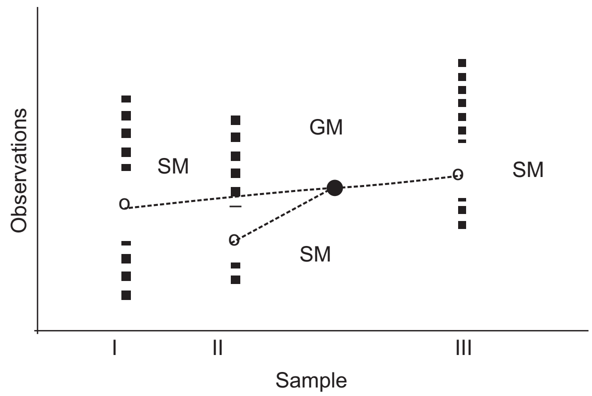


FIGURE 9.2 ANOVA decomposition. *GM*, grand mean; *SM*, sample mean.

As mentioned earlier, *SST* estimates the variation among the $\mu_i' s$, and hence, if all the $\mu_i' s$ were equal, the $\overline{T_i} s$ would be similar and the *SST* would be small. It can be verified that the unbiased estimator of σ^2 based on $(n_1 + n_2 + \cdot \cdot \cdot + n_k - k)$ degrees of freedom is:

$$S^2 = MSE = \frac{SSE}{(n_1 + n_2 + \cdots + n_k - k)}$$

$$= \frac{SSE}{N - k}.$$

Note that the quantity *MSE* is a measure of variability within the groups. If there were only one group with n observations, then the *MSE* would be nothing but the sample variance, s^2. The fact that ANOVA deals simultaneously with all k groups can be seen by rewriting MSE in the following form:

$$MSE = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1)}.$$

The *mean square for treatments* with $(k - 1)$ degrees of freedom is:

$$MST = \frac{SST}{k - 1}.$$

The *MST* is a measure of the variability between the sample means of the groups. We now summarize the ANOVA hypothesis-testing method for two or more populations.

One-way analysis of variance for $k \geq 2$ populations

We test:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{versus}$$
$$H_a: \text{At least two of the } \mu_i' \text{s are different.}$$

When H_0 is true, we have:

$$E(MST) = E(MSE).$$

The greater the differences among the $\mu' s$, the larger the E(MST) will be relative to E(MSE).

Test statistic:

$$F = \frac{MST}{MSE}.$$

Rejection region is:

$$RR: F > F_\alpha$$

with $\upsilon_1 = (k - 1)$ numerator degrees of freedom and $\upsilon_2 = \sum_{i=1}^{k} n_i - k = N - k$ denominator degrees of freedom, where $N = \sum_{i=1}^{k} n_i$.

Assumptions: The observations $Y_{ij}' s$ are assumed to be independent and normally distributed with mean μ_i, $i = 1, 2, \ldots, k$, and variance σ^2.

Now we give a five-step computational procedure that we could follow for the ANOVA for the completely randomized design.

One-way analysis of variance procedure for $k \geq 2$ populations

We test:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k \quad \text{versus}$$
$$H_a: \text{At least two of the } \mu_i's \text{ are different.}$$

1. Compute:

$$T_i = \sum_{j=1}^{n_i} y_{ij},\ T = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij},\ \text{and} \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2,$$

$$CM = \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}\right)^2}{N} = \frac{T^2}{N},\ \text{where } N = \sum_{i=1}^{k} n_i,$$

$$\overline{T}_i = \frac{T_i}{n_i},$$

and

$$Total\ SS = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - CM.$$

2. Compute the sum of squares between samples (treatments),

$$SST = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - CM,$$

and the sum of squares within samples,

$$SSE = Total\ SS - SST.$$

Let

$$MST = \frac{SST}{k-1},$$

and

$$MSE = \frac{SSE}{n-k}.$$

3. Compute the test statistic:

$$F = \frac{MST}{MSE}.$$

4. For a given α, find the rejection region as:

$$RR: F > F_\alpha$$

with $\upsilon_1 = (k-1)$ numerator degrees of freedom and $\upsilon_2 = \left(\sum_{i=1}^{k} n_i\right) - k = N - k$ denominator degrees of freedom, where $N = \sum_{i=1}^{k} n_i$.

5. **Conclusion:** If the test statistic F falls in the rejection region, conclude that the sample evidence supports the alternative hypothesis that at least one pair of the means is indeed different for the k treatments and all are not equal.

Assumptions: The samples are randomly selected from the k populations in an independent manner. The populations are assumed to be normally distributed with equal variances σ^2 and means $\mu_1, \ldots, \mu_k$.

Even though the completely randomized design is extremely easy to construct and the calculations described above are relatively easy, the homogeneousness of the treatments is crucial. Any extraneous sources of variability will make it more difficult to detect differences among treatment means due to inflation of the error term.

9.3.1 The p-value approach

Note that if we are using statistical software packages, the p-value approach can be used for the testing. Just compare the p value and α to arrive at a conclusion. Refer to the computer examples in Section 9.7.

The following example illustrates the ANOVA procedure.

EXAMPLE 9.3.1

We are given three random samples as shown in Table 9.2 that represent test scores from three classes of statistics taught by three different instructors and are independently sampled from each class. Assume that the three different populations are normal with equal variances.

TABLE 9.2 Test Scores for Three Classes.

Sample 1	Sample 2	Sample 3
64	56	81
84	74	92
75	69	84
77		
80		

At the $\alpha = 0.05$ level of significance, test for equality of the population means.

Solution
We test:

$$H_0: \mu_1 = \mu_2 = \mu_3 \text{ versus } H_a: \text{At least two of the } \mu's \text{ are different.}$$

Here, $k = 3$, $n_1 = 5$, $n_2 = 3$, *and* $N = n_1 + n_2 + n_3 = 11$.
Also,

T_i	380	199	257
n_i	5	3	3
$\overline{T}_i$	76	66.33	85.67

Clearly, the sample means are different. The question we are going to answer is, Is this difference due to just chance, or is it due to a real difference caused by different teaching styles? For this, we now compute the following:

$$CM = \frac{\left(\sum_i \sum_j y_{ij}\right)^2}{N} = \frac{(836)^2}{11} = 63,536,$$

$$Total\ SS = \sum_i \sum_j y_{ij}^2 - CM$$

$$= 64,560 - 63,536 = 1024,$$

$$SST = \sum_i \frac{T_i^2}{n_i} - CM$$

$$= \frac{(380)^2}{5} + \frac{(199)^2}{3} + \frac{(257)^2}{3} - CM$$

$$= 64,096.66 - 63,536 = 560.66,$$

and

$$SSE = Total\ SS - SST$$

$$= 1024 - 560.66 = 463.34.$$

Hence,

$$MST = \frac{SST}{k-1} = \frac{560.66}{2} = 280.33,$$

and

$$MSE = \frac{SSE}{N-k} = \frac{463.34}{8} = 57.9175.$$

The test statistic is:

$$F = \frac{MST}{MSE} = \frac{280.33}{57.9175} = 4.84.$$

From the F *table,* $F_{0.05,2,8} = 4.46$.
Therefore, the rejection region is given by:

$$RR: F > 4.46.$$

Decision: Because the observed value of $F = 4.84$ *falls in the rejection region, we do reject* H_0 *and conclude that there is sufficient evidence to indicate a difference in the true means.*

If we want the p *value, we can see from the* F *table that* $0.025 < p$ *value* < 0.05, *indicating the rejection of the null hypothesis with* $\alpha = 0.05$. *Using statistical software packages, we can get the exact* p *value.*

The calculations obtained in analyzing the total sum of squares into its components are usually summarized by the *analysis-of-variance table* (ANOVA table), given in Table 9.4.

Sometimes, one may also add a column for the p value, $P(F_{k-1,\ n-k} \geq \text{observed } F)$, in the ANOVA table.

For the previous example, we can summarize the computations in the ANOVA table shown in Table 9.3.

TABLE 9.3 ANOVA Table.

Source of variation	Degrees of freedom	Sum of squares	Mean squares	*F*-statistic
Treatments	$k-1$	$SST = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CM$	$MST = \frac{SST}{k-1}$	$\frac{MST}{MSE}$
Error	$N-k$	$SSE = Total\ SS - SST$	$MSE = \frac{SSE}{N-k}$	
Total	$N-1$	$Total\ SS = \sum_{i=1}^{k} \sum_{i=1}^{n_i} \left(y_{ij} - \bar{y}\right)^2$		

TABLE 9.4 ANOVA Table for Test Scores.

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F*-statistic	*p* value
Treatments	2	560.66	280.33	4.84	0.042
Error	8	463.34	57.917		
Total	10	1024			

9.3.2 Testing the assumptions for one-way analysis of variance

The randomness assumption could be tested using the Wald−Wolfowitz test (see Project 12B). The assumption of independence of the samples is hard to test without knowing how the data are collected and should be implemented during collection of data in the design stage. Normality can be tested (this should be performed separately for each sample, not for the total data set) using probability plots or other tests such as the chi-square goodness-of-fit test. ANOVA is fairly robust against violation of this assumption if the sample sizes are equal. Also, if the sample sizes are fairly large, the central limit theorem helps. The presence of outliers is likely to increase the sample variance, thus decreasing the value of the *F*-statistic for ANOVA, which will result in a lower power of the test. Box plots or probability plots could be used to identify the outliers. If the normality test fails, transforming the data (see Section 14.4.2) or a nonparametric test such as the Kruskal−Wallis test described in Section 12.5.1 may be more appropriate. If the sizes of all the samples are equal, ANOVA is mostly robust for violation of homogeneity of the variances. A rule of thumb used for robustness for this condition is that the ratio of sample variance of the largest sample variance s^2 to the smallest sample variance s^2 should be no more than 3:1. Another popular rule of thumb used in one-way ANOVA to verify the requirement of equality of variances is that the largest sample standard deviation not be larger than two times the smallest sample standard deviation. Graphically, representing side-by-side box plots of the samples can also reveal a lack of homogeneity of variances if some box plots are much longer than others (see Fig. 9.3E). For a significance test on the homogeneity of variances (Levene's test), refer to Section 14.4.3. If these tests reveal that the variances are different, then the populations are different, despite what ANOVA concludes about differences of the means. But this itself is significant, because it shows that the treatments had an effect.

EXAMPLE 9.3.2

In order to study the effect of automobile size on noise pollution, the following data are randomly chosen from the air pollution data (source: A.Y. Lewin and M.F. Shakun, *Policy Sciences: Methodology and Cases*, Pergamon Press, 1976, p. 313). The automobiles are categorized as small, medium, and large, and noise level readings (in decibels) are given in Table 9.5.

TABLE 9.5 **Size of Automobile and Noise Level (Decibels).**

	Size of automobile		
	Small	Medium	Large
Noise level (dB)	820	840	785
	820	825	775
	825	815	770
	835	855	760
	825	840	770

At the $\alpha = 0.05$ level of significance, test for equality of population mean noise levels for different sizes of the automobiles. Comment on the assumptions.

Solution

Let μ_1, μ_2, and μ_3 be population mean noise levels for small, medium, and large automobiles, respectively. First, we test for the assumptions. Using Minitab, run tests for each of the samples; we can justify the assumption of randomness of the sample values. A normality test for each column gives the graphs shown in Figs. 9.3A–9.3C, through which we can reasonably assume normality. Because the sample sizes are equal, we will use the one-way ANOVA method to analyze these data.

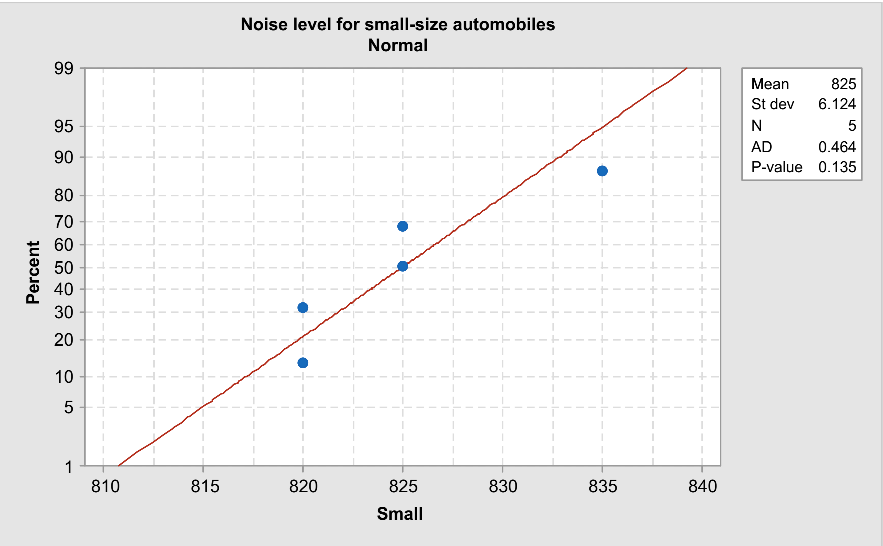


FIGURE 9.3A Normal plot for noise level of small automobiles.

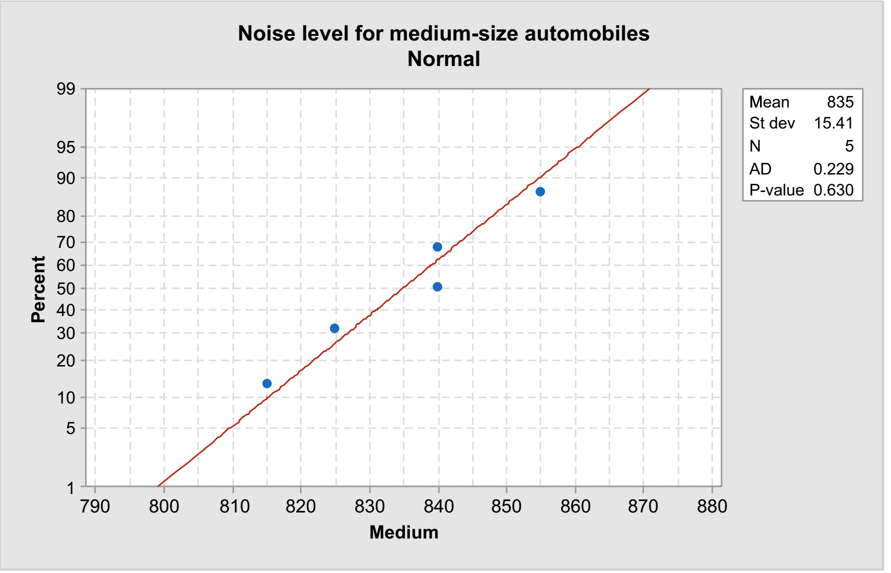


FIGURE 9.3B Normal plot for noise level of medium-sized automobiles.

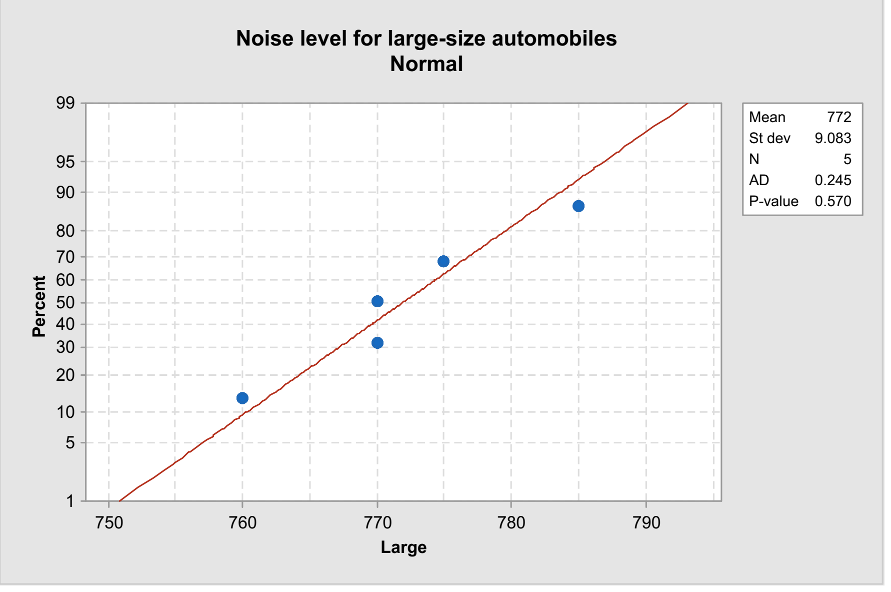


FIGURE 9.3C Normal plot for noise level of large automobiles.

Fig. 9.3D indicates that the relative positions of the sample means are different, and Fig. 9.3E (Minitab steps for creating side-by-side box plots are given at the end of Example 9.7.1) gives an indication of within-group variations; perhaps the group 2 (medium size) variance is larger. Now, we will do the analytic testing.

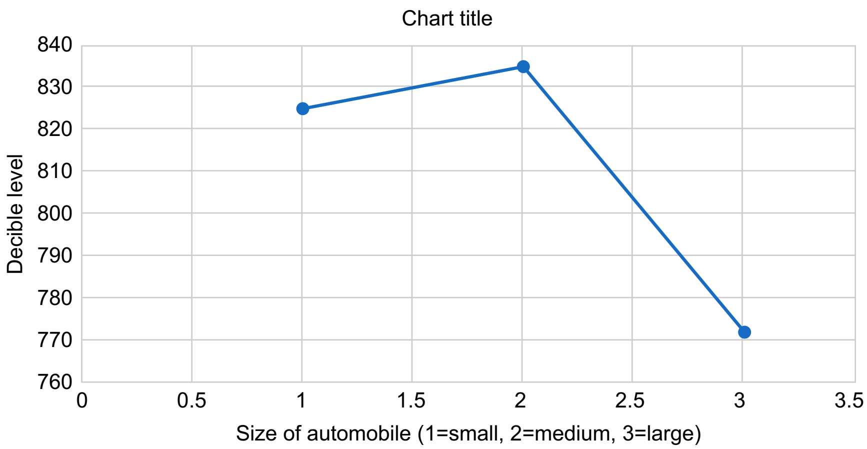


FIGURE 9.3D Mean decibel levels for three sizes of automobiles.

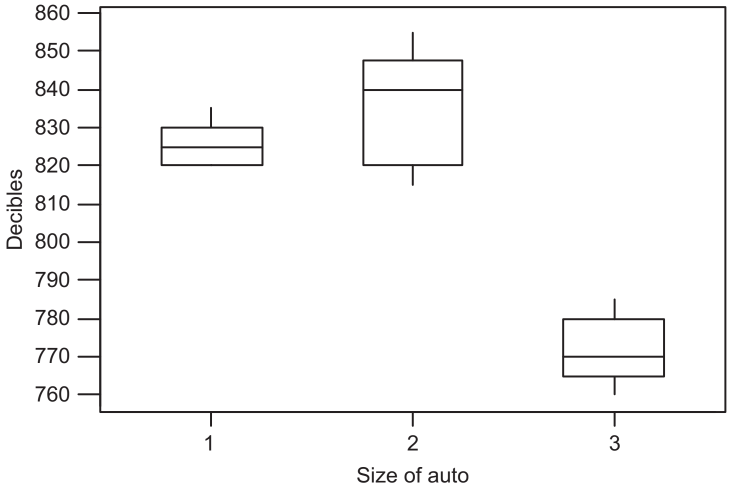


FIGURE 9.3E Side-by-side box plots for decibel levels for three sizes of automobiles.

We test:

$$H_0: \mu_1 = \mu_2 = \mu_3 \text{ } \textit{versus} \text{ } H_a: \textit{At least two of the } \mu'\textit{s are different}.$$

Here, $k = 3$, $n_1 = 5$, $n_2 = 5$, $n_3 = 5$, *and* $N = n_1 + n_2 + n_3 = 15$.

Also,

T_i	4125	4175	3860
n_i	5	5	5
$\overline{T}_i$	825	835	772

In the following calculations, for convenience we will approximate all values to the nearest integer:

$$CM = \frac{\left(\sum_i \sum_j y_{ij}\right)^2}{N} = \frac{(12,160)^2}{15} = 9,857,707,$$

$$Total\ SS = \sum_i \sum_j y_{ij}^2 - CM$$

$$= 12,893,$$

$$SST = \sum_i \frac{T_i^2}{n_i} - CM$$

$$= 11,463,$$

and

$$SSE = Total\ SS - SST$$

$$= 1430.$$

Hence,

$$MST = \frac{SST}{k-1} = \frac{11,463}{2} = 5732,$$

and

$$MSE = \frac{SSE}{N-k} = \frac{1430}{12} = 119.$$

The test statistic is:

$$F = \frac{MST}{MSE} = \frac{5732}{119} = 48.10.$$

From the table, we get $F_{0.05,2,12} = 3.89$. *Because the test statistic falls in the rejection region, we reject at* $\alpha = 0.05$ *the null hypothesis that the mean noise levels are the same. We conclude that the size of the automobile does affect the mean noise level.*

It should be noted that the alternative hypothesis H_a in this section covers a wide range of situations, from the case where all but one of the population means are equal to the case where they are all different. Hence, with such an alternative, if the samples lead us to reject the null hypothesis, we are left with a lot of unsettled questions about the means of the k populations. These are called *post hoc* testing. This problem of multiple comparisons is the topic of Section 9.8.

9.3.3 Model for one-way analysis of variance (optional)

We conclude this section by presenting the classical model for one-way ANOVA. Because the variables Y_{ij} are random samples from normal populations with $E(Y_{ij}) = \mu_i$ and with common variance $Var(Y_{ij}) = \sigma^2$, for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$, we can write a model as:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i$$

where the error terms ε_{ij} are independent normally distributed random variables with $E(\varepsilon_{ij}) = 0$ and $Var(\varepsilon_{ij}) = \sigma^2$. Let $\alpha_i = \mu_i - \mu$ be the difference of μ_i (ith population mean) from the grand mean μ. Then α_i can be considered as the ith treatment effect. Note that the α_i values are nonrandom. Because $\mu = \Sigma_i (n_i \mu_i / N)$, it follows that $\sum_{i=1}^k \alpha_i = 0$. This will result in the following classical model for one-way layout:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i.$$

With this representation, the test H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ reduces to testing the null hypothesis that there is no treatment effect, H_0: $\alpha_i = 0$, for $i = 1, \ldots, k$.

Exercises 9.3

9.3.1. In an effort to investigate the premium charged by insurance companies for auto insurance, an agency randomly selects a few drivers who are insured by one of three different companies. These individuals have similar cars, driving records, and levels of coverage. Table 9.6 gives the premiums paid per 6 months by these drivers with the three companies.

TABLE 9.6 Auto Insurance Premiums.

Company I	Company II	Company III
396	348	378
438	360	330
336	522	294
318	474	432

(a) Construct an ANOVA table and interpret the results.
(b) Using the 5% significance level, test the null hypothesis that the mean auto insurance premium paid per 6 months by all drivers insured for each of these companies is the same. Assume that the conditions of completely randomized design are met.

9.3.2. Three classes in elementary statistics are taught by three different persons: a regular faculty member, a graduate teaching assistant, and an adjunct from outside the university. At the end of the semester, each student is given a standardized test. Five students are randomly picked from each of these classes, and their scores are as shown in Table 9.7.

TABLE 9.7 Test Scores by Instructor Type.

Faculty	Teaching assistant	Adjunct
93	88	86
61	90	56
87	76	73
75	82	90
92	58	47

(a) Construct an ANOVA table and interpret your results.
(b) Test at the 0.05 level whether there is a difference between the mean scores for the three persons teaching. Assume that the conditions of completely randomized design are met.

9.3.3. Let $n_1 = n_2 = \ldots = n_k = n'$. Show that:

$$\sum_{i=1}^{k}\sum_{j=1}^{n'}\left(y_{ij}-\overline{y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n'}\left(y_{ij}-\overline{T}_i\right)^2 + n\sum_{i=1}^{k}\left(\overline{T}_i-\overline{y}\right)^2.$$

9.3.4. For the sum of squares for treatment:

$$SST = \sum_{i=1}^{k} n_i\left(\overline{T}_i-\overline{y}\right)^2,$$

show that:

$$E(SST) = (k-1)\sigma^2 + \sum_{i=1}^{k} n_i(\mu_i-\mu)^2$$

where $\mu = \frac{1}{N}\Sigma_{i=1}^{k} n_i\mu_i$.
[This exercise shows that the expected value of SST increases as the differences among the $\mu_i's$ increase.]

9.3.5 (a) Show that:

$$SSE = \sum_{i=1}^{k}(n_i - 1)S_i^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{T}_i\right)^2,$$

where $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{T}_i\right)^2$ provides an independent, unbiased estimator for σ^2 in each of the k samples.

(b) Show that SSE/σ^2 has a chi-square distribution with $N - k$ degrees of freedom, where $N = \sum_{i=1}^{k} n_i$.

9.3.6. Let each observation in a set of k independent random samples be normally distributed with means $\mu_1, \ldots, \mu_k$ and common variance σ^2. If $H_0 = \mu_1 = \mu_2 = \ldots = \mu_k$ is true, show that:

$$F = \frac{SST/(k-1)}{SSE/(n-k)} = \frac{MST}{MSE},$$

has an F distribution with $k - 1$ numerator and $n - k$ denominator degrees of freedom.

9.3.7. The management of a grocery store observes various employees for work productivity. Table 9.8 gives the number of customers served by each of its four checkout lanes per hour.

TABLE 9.8 Number of Customers Served by Different Employees.

Lane 1	Lane 2	Lane 3	Lane 4
16	11	8	21
18	14	12	16
22	10	17	17
21	10	10	23
15	14	13	17
	10	15	

(a) Construct an ANOVA table and interpret the results. Indicate any assumptions that were necessary.

(b) Test whether there is a difference between the mean numbers of customers served by the four employees at the 0.05 level. Assume that the conditions of completely randomized design are met.

9.3.8. Table 9.9 represents immunoglobulin levels (with each observation being the IgA immunoglobulin level measured in international units) of children under 10 years of age of a particular group. The children are grouped as follows: group A, ages 1 to less than 3; group B, ages 3 to less than 6; group C, ages 6 to less than 8; and group D, ages 8 to less than 9. Test whether there is a difference between the means for each of the age groups. Use $\alpha = 0.05$. Interpret your results and state any assumptions that were necessary to solve the problem.

TABLE 9.9 Immunoglobulin Level by Age Group.

A	35	8	12	19	56	64	75	25		
B	31	79	60	45	39	44	45	62	20	66
C	74	56	77	35	95	81	28			
D	80	42	48	69	95	40	86	79	51	

9.3.9. Table 9.10 gives rental and homeowner vacancy rates by US region (source: US Census Bureau) for 5 years. Test at the 0.01 level whether the true rental and homeowner vacancy rates by area are the same for all 5 years. Interpret your results and state any assumptions that were necessary to perform the analysis.

TABLE 9.10 Rental Vacancy by Region.

Rental units	1995	1996	1997	1998	1999
Northeast	7.2	7.4	6.7	6.7	6.3
Midwest	7.2	7.9	8.0	7.9	8.6
South	8.3	8.6	9.1	9.6	10.3
West	7.5	7.2	6.6	6.7	6.2

9.3.10. Table 9.11 gives lower limits of income (approximated to the nearest \$1000 and calculated as of March of the following year) of the top 5% of US households by race from 1994 to 1998 (source: US Census Bureau).
Test at the 0.05 level whether the true lower limits of income for the top 5% of US households for each race are the same for all 5 years.

TABLE 9.11 Lower Limits of Income of Top 5% by Race.

Race	Year				
	1994	1995	1996	1997	1998
All races	110	113	120	127	132
White	113	117	123	130	136
Black	81	80	85	87	94
Hispanic	82	80	86	93	98

9.3.11. Table 9.12 gives mean serum cholesterol levels (given in milligrams per deciliter) by race and age for the adult population in the United States between 1978 and 1980.
(Source: Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, *Arch. Intern. Med.* 148, January 1988.)
Test at the 0.01 level whether the true mean cholesterol levels for the adult population in the United States between 1978 and 1980 are the same.

TABLE 9.12 Mean Serum Level by Race and Age.

	Age					
Race	20–24	25–34	35–44	45–54	55–64	65–74
All races	180	199	217	227	229	221
White	180	199	217	227	230	222
Black	171	199	218	229	223	217

9.4 Two-way analysis of variance, randomized complete block design

A *randomized block design*, or the *two-way ANOVA*, consists of b blocks of k experimental units each. In many cases we may be required to measure response at combinations of levels of two or more factors considered simultaneously. For example, we might be interested in gas mileage per gallon among four different makes of cars for both in-city and highway driving, or in examining weight loss comparing five different diet programs among whites, African Americans, Hispanics, and Asians according to their gender. In studies involving various factors, the effect of each factor on the response variable may be analyzed using one-way classification. However, such an analysis will not be efficient with respect to time, effort, and cost. Also, such a procedure would give no knowledge about the likely interactions that may exist among different factors. In such cases, the two-way ANOVA is an appropriate statistical method to use.

In a randomized block design, the treatments are randomly assigned to the units in each block, with each treatment appearing exactly once in every block (that is, there is no interaction between factors). Thus, the total number of observations obtained in a randomized block design is $n = bk$. The purpose of subdividing experiments into blocks is to eliminate as much variability as possible, that is, to reduce the experimental error or the variability due to extraneous causes. Refer to Section 9.2.3 for a procedure to obtain completely randomized block design. The goal of such an experiment is to test the equality of levels for the treatment effect. Sometimes, it may also be of interest to test for a difference among blocks. We proceed to give a formal statistical model for the completely randomized block design.

For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, b$, let $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, where Y_{ij} is the observation on treatment i in block j, μ is the overall mean, α_i is the nonrandom effect of treatment i, β_j is the nonrandom effect of block j, and ε_{ij} are the random error terms such that ε_{ij} are independent normally distributed random variables with $E(\varepsilon_{ij}) = 0$ and $Var(\varepsilon_{ij}) = \sigma^2$. In this case, $\Sigma\alpha_i = 0$, and $\Sigma\beta_j = 0$.

The ANOVA for a randomized block design proceeds similar to that for a completely randomized design, the main difference being that the total sum of squares of deviations of the response measurements from their means may be partitioned into three parts: the sum of squares of blocks (SSB), treatments (SST), and error (SSE).

Let $B_j = \sum_{i=1}^{k} y_{ij}$ and $\overline{B}_j$ denote, respectively, the total sum and mean of all observations in block j. Represent the total for all observations receiving treatment i by $T_i = \sum_{j=1}^{b} y_{ij}$, and mean and $\overline{T}_i$, respectively. Let

$$\bar{y} = \text{average of } n = bk \text{ observations}$$

$$= \frac{1}{n}\sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij}$$

and

$$CM = \frac{1}{n}(\text{total of all observations})^2$$

$$= \frac{1}{n}\left(\sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij}\right)^2.$$

For convenience, we can represent the two-way classification as in Table 9.13.

Note that from the table we can obtain $\sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij} = \sum_{j=1}^{b} B_j$. Hence, $CM = (1/n)\left(\sum_{j=1}^{b} B_j\right)^2$.

TABLE 9.13 Two-way Classification.

	Blocks							
	1	**2**	…	j	…	b	**Total** T_i	**Mean** $\overline{T}_i$
Treatment 1	y_{11}	y_{12}	…	y_{1j}	…	y_{1b}	T_1	$\overline{T}_1$
Treatment 2	y_{21}	y_{22}	…	y_{2j}	…	y_{2b}	T_2	$\overline{T}_2$
.	.		.					
.	.		.					
.	.		.					
Treatment i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ib}	T_i	$\overline{T}_i$
.					.			
.					.			
.					.			
Treatment k	y_{k1}	y_{k2}	…	y_{kj}	…	y_{kb}	T_k	$\overline{T}_k$
Total B_j	B_1	B_2	…	B_j	…	B_b		
Mean $\overline{B}_j$	$\overline{B}_1$	$\overline{B}_2$	…	$\overline{B}_j$	…	$\overline{B}_b$		$\bar{y}$

Then, for a randomized block design with b blocks and k treatments, we need to compute the following sums of squares. They are:

$$Total\ SS = SSB + SST + SSE$$

$$= \sum_{j=1}^{b} \sum_{i=1}^{k} \left(y_{ij} - \overline{y}\right)^2 = \sum_{j=1}^{b} \sum_{i=1}^{k} y_{ij}^2 - CM,$$

$$SSB = k \sum_{j=1}^{b} \left(\overline{B}_j - \overline{y}\right)^2 = \frac{\sum_{j=1}^{b} B_j^2}{k} - CM,$$

$$SSB = b \sum_{i=1}^{k} \left(\overline{T}_i - \overline{y}\right)^2 = \frac{\sum_{i=1}^{k} T_i^2}{b} - CM,$$

and

$$SSE = Total\ SS - SSB - SST.$$

We define:

$$MSB = \frac{SSB}{b-1},$$

$$MST = \frac{SST}{k-1},$$

and

$$MSE = \frac{SSE}{n-b-k+1}.$$

The ANOVA for the randomized block design is presented in Table 9.14. The column corresponding to d.f. represents the degrees of freedom associated with each sum of squares. MS denotes the mean square.

TABLE 9.14 ANOVA Table for Randomized Block Design.

Source	d.f.	SS	MS
Blocks	$b - 1$	SSB	$\frac{SSB}{b-1}$
Treatments	$k - 1$	SST	$\frac{SST}{k-1}$
Error	$(b-1)(k-1) = n - b - k + 1$	SSE	$\frac{SST}{n-b-k+1}$
Total	$n - 1$	$Total\ SS$	

To test the null hypothesis that there is no difference in treatment means, that is, to test:

$$H_0: \alpha_i = 0, \quad i = 1, \ldots, k \text{ versus } H_a: \text{Not all } \alpha_i' s \text{ are zero.}$$

we use the F-statistic,

$$F = \frac{MST}{MSE},$$

and reject H_0 if $F > F_\alpha$ based on $(k-1)$ numerator and $(n-b-k+1)$ denominator degrees of freedom.

Although blocking lowers the experimental error, it also furnishes a chance to see whether evidence exists to indicate a difference in the mean response for blocks. In this case we will be testing the hypothesis:

$$H_0: \beta_j = 0, \quad j = 1, \ldots, b \text{ versus } H_a: \text{Not all } \beta_j' s \text{ are zero.}$$

Under the assumption that there is no difference in the mean response for blocks, MSB provides an unbiased estimator for σ^2 based on $(b-1)$ degrees of freedom. If there is a real difference that exists among block means, MSB will be larger in comparison with MSE and

$$F = \frac{MSB}{MSE},$$

will be used as a test statistic. The rejection region in this case will be $F > F_\alpha$ based on $(b-1)$ numerator and $(n-b-k+1)$ denominator degrees of freedom.

We now summarize the foregoing methodology in a step-by-step computational procedure. For a reasonable data set size, we could use scientific calculators for handling the ANOVA calculations. For larger data sets, the use of statistical software packages is recommended.

Computational procedure for randomized block design

1. Calculate the following quantities:
   (i) Sum the observations for each row to form row totals:
   $$T_1, T_2, \ldots, T_k, \text{ where } T_i = \sum_{j=1}^{b} y_{ij}.$$
   (ii) Sum the observations for each column to form column totals:
   $$B_1, B_2, \ldots, B_b, \text{ where } B_j = \sum_{i=1}^{k} y_{ij}.$$
   (iii) Find the sum of all observations:
   $$\sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij} = \sum_{j=1}^{b} B_j.$$
2. Calculate the following quantities:
   (i) Square the sum of the totals for each column and divide it by $n = bk$ to obtain:
   $$CM = \frac{1}{n}\left(\sum_{j=1}^{b} B_j\right)^2.$$
   (ii) Find the sum of squares of the totals of each column and divide it by k to obtain:
   $$\frac{1}{k}\sum_{j=1}^{b} B_j^2$$
   and
   $$SSB = \frac{\sum_{j=1}^{b} B_j^2}{k} - CM \quad \text{and} \quad MSB = \frac{SSB}{b-1}.$$
   (iii) Find the sum of squares of the totals of each row and divide it by b to obtain:
   $$\frac{\sum_{i=1}^{k} T_i^2}{b},$$
   and
   $$SST = \frac{\sum_{i=1}^{k} T_j^2}{b} - CM, \quad \text{and} \quad MSB = \frac{SST}{k-1}.$$
   (iv) Find the sum of squares of individual observations:
   $$\sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij}^2.$$
   Also compute:
   $$Total\ SS = \sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij}^2 - CM.$$
   (v) Using (ii), (iii), and (iv), find:
   $$SSE = Total\ SS - SSB - SST \quad \text{and} \quad MSE = \frac{SSE}{n-b-k+1}.$$
3. To test the null hypothesis that there is no difference in treatment means:
   (i) Compute the F-statistic,
   $$F = \frac{MST}{MSE}.$$
   (ii) From the F table, find the value of $F_{\alpha,\upsilon_1,\upsilon_2}$, where $\upsilon_1 = (k-1)$ is the numerator and $\upsilon_2 = (n-b-k+1)$ is the denominator degrees of freedom.
   (iii) **Decision:** Reject H_0 if $F > F_{\alpha,\upsilon_1,\upsilon_2}$ and conclude that there is evidence to conclude that there is a difference in treatment means at level α.
4. To test the null hypothesis that there is no difference in the mean response for blocks:
   (i) Compute the F-statistic,
   $$F = \frac{MSB}{MSE}.$$
   (ii) From the F table, find the value of $F_{\alpha,\upsilon_1,\upsilon 2}$, where $\upsilon_1 = (b-1)$ is the numerator and $\upsilon_2 = (n-b-k+1)$ is the denominator degrees of freedom.

Computational procedure for randomized block design—cont'd

(iii) Decision: Reject H_0 if $F > F_{\alpha,v_1,v2}$ and conclude that there is evidence to conclude there is a difference in the mean response for blocks at level α.

Assumptions: The samples are randomly selected in an independent manner from $n = bk$ populations. The populations are assumed to be normally distributed with equal variances σ^2. Also, there are no interactions between the variables (two factors).

We have already discussed the assumptions and how to verify those assumptions in one-way analysis. The only new assumption in the randomized blocked design is about the interactions. One of the ways to verify the assumption of no interaction is to plot the observed values against the sample number. If there is no interaction, the line segments (one for each block) will be parallel or nearly parallel; see Fig. 9.2. If the lines are not approximately parallel, then there is likely to be interaction between blocks and treatments. In the presence of interactions, the analysis of this section needs to be modified. For details on those procedures, refer to more specialized books on ANOVA methods.

We illustrate the randomized block design procedure with the following example.

EXAMPLE 9.4.1

A furniture company wants to know whether there are differences in stain resistance among the four chemicals used to treat three different fabrics. Table 9.15 shows the yields of resistance to stain (a low value indicates good stain resistance).

At the $\alpha = 0.05$ level of significance, is there evidence to conclude that there is a difference in mean resistance among the four chemicals? Is there any difference in the mean resistance among the materials? Give bounds for the p values in each case.

TABLE 9.15 Stain Resistance by Chemicals and by Fabric Types.

Chemical	Material			
	I	II	III	Total
C_1	3	7	6	16
C_2	9	11	8	28
C_3	2	5	7	14
C_4	7	9	8	24
Total	21	32	29	82

Solution

Here, $T_1 = 16$, $T_2 = 28$, $T_3 = 14$, *and* $T_4 = 24$. *Also,* $B_1 = 21$, $B_2 = 32$, *and* $B_3 = 29$. *In addition,* $b = 3$, $k = 4$, *and* $n = bk = 12$. *Now:*

$$CM = \frac{1}{n}\left(\sum_{j=1}^{b} B_j\right)^2 = \frac{1}{12}(82)^2 = 560.3333.$$

We can compute the following quantities:

$$SSB = \frac{\sum_{j=1}^{b} B_j^2}{k} - CM = \frac{2306}{4} - 560.3333 = 16.1667,$$

$$MSB = \frac{SSB}{b-1} = \frac{16.1667}{2} = 8.0834,$$

$$SST = \frac{\sum_{i=1}^{k} T_i^2}{b} - CM = \frac{1812}{3} - 560.3333 = 43.6667,$$

and

$$MST = \frac{SST}{k-1} = \frac{43.6667}{3} = 14.5556.$$

We have $\sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij}^2 = 632$. Thus,

$$Total\ SS = \sum_{j=1}^{b}\sum_{i=1}^{k} y_{ij}^2 - CM = 632 - 560.3333 = 71.666,$$

$$SSE = Total\ SS - SSB - SST = 71.6667 - 16.1667 - 43.6667$$

$$= 11.8333,$$

and

$$MSE = \frac{SSE}{n-b-k+1} = \frac{11.8333}{6} = 1.9722.$$

The F-*statistic is:*

$$F = \frac{MST}{MSE} = \frac{14.5556}{1.9722} = 7.3804.$$

From the F-*table,* $F_{0.05,3,6} = 4.76$. *Because the observed value* $F = 7.3804 > 4.76$, *we reject the null hypothesis and conclude that there is a difference in mean resistance among the four chemicals. Because the* F-*value falls between* $\alpha = 0.025$ *and* $\alpha = 0.01$, *the* p *value falls between* 0.01 *and* 0.025. *To test for the difference in the mean resistance among the materials,*

$$F = \frac{MSB}{MSE} = \frac{8.0834}{1.9722} = 4.0987.$$

From the F *table,* $F_{0.05,2,6} = 5.14$. *Because the observed value of* $F = 4.098 < 5.14$, *we conclude that there is no difference in the mean resistance among the materials. Because the* F *value falls between* $\alpha = 0.10$ *and* 0.05, *the* p *value falls between* 0.05 *and* 0.9.

Exercises 9.4

9.4.1. Show that:

$$\sum_{j=1}^{b}\sum_{i=1}^{k}\left(y_{ij} - \overline{y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{b}\left(y_{ij} - \overline{T}_i - \overline{B}_j - \overline{y}\right)^2$$

$$+ b\sum_{i=1}^{k}\left(\overline{T}_i - \overline{y}\right)^2 + k\sum_{j=1}^{b}\left(\overline{B}_j - \overline{y}\right)^2.$$

Hint: Use the identity $\left[y_{ij} - \overline{y} = \left(y_{ij} - \overline{T}_i - \overline{B}_j - \overline{y}\right) + \left(\overline{T}_i - \overline{y}\right) + \left(\overline{B}_j - \overline{y}\right)\right]$

9.4.2. Show the following:

(a) $E(MSE) = \sigma^2$,

(b) $E(MSB) = \frac{k}{b-1}\sum_{j=1}^{b} B_j^2 + \sigma^2$,

(c) $E(MST) = \frac{b}{k-1}\sum_{i=1}^{k} \tau_i^2 + \sigma^2$.

9.4.3. The least-square estimators of the parameters μ, τ_i, and β_j are obtained by minimizing the sum of squares:

$$W = \sum_{i=1}^{k}\sum_{j=1}^{b}\left(y_{ij} - \mu - \tau_i - \beta_j\right)^2,$$

with respect to μ, τ_i, and β_j, subject to the restrictions $\sum_{i=1}^{k}\tau_i = \sum_{j=1}^{b}\beta_j = 0$. Show that the resulting estimators are:

$$\widehat{\mu} = \overline{y},$$

$$\widehat{\tau}_i = \overline{T}_i - \overline{y}, i = 1, 2, \ldots, k,$$

and

$$\widehat{\beta}_j = \overline{B}_j - \overline{y}, \quad j = 1, \ldots, b.$$

9.4.4 To test the wear on four hyperalloys, a test piece of each alloy was extracted from each of the three positions of a test machine. The reduction of weight in milligrams due to wear was determined on each piece, and the data are given in Table 9.16.

TABLE 9.16 **Loss of weight due to wear testing of four materials (in mg).**

	Position		
Type of alloy	**1**	**2**	**3**
1	241	270	274
2	195	241	218
3	235	273	230
4	234	236	227

At $\alpha = 0.05$, test the following hypotheses, regarding the positions as blocks:

(a) There is no difference in average wear for each material.
(b) There is no difference in average wear for each position.
(c) Interpret your final result and state any assumptions that were necessary to solve the problem.

9.4.5. Using the data of Exercise 9.3.10, test at the 0.05 level that the true income lower limits of the top 5% of US households for each race are the same for all 5 years. Also, test at the 0.05 level that the true income lower limits of the top 5% of US households for each year between 1994 and 1998 are the same.

9.4.6. Using the data of Exercise 9.3.11, test at the 0.01 level that the true mean cholesterol levels for all races in the United States during 1978–80 are the same. Also, test at the 0.01 level that the true mean cholesterol levels for all ages in the United States during 1978–80 are the same.

9.4.7. To see the effect of hours of sleep on tests of different skill categories (vocabulary, reasoning, and arithmetic), tests consisting of 20 questions in each category were given to 16 students, in groups of four, based on the hours of sleep they had on the previous night. Each right answer is given one point. Table 9.17 gives the cumulative scores of each group of four students in each category.

TABLE 9.17 **Effect of Sleep on Test Scores by Skill Categories**

Hours of sleep	**Category**		
	Vocabulary	**Reasoning**	**Arithmetic**
0	44	33	35
4	54	38	18
6	48	42	43
8	55	52	50

Test at the 0.05 level whether the true mean performance for different hours of sleep is the same. Also, test at the 0.05 level whether the true mean performance for each category of the test is the same.

9.5 Multiple comparisons

The ANOVA procedures that we have used so far showed whether differences among several means are significant. However, if the equality of means is rejected, the F-test did not pinpoint for us which of the given means or groups of means differs significantly from another given mean or group of means. With ANOVA, when the null hypothesis of equality of means is rejected, the problem is to see whether there is some way to follow up (post hoc) this initial test, H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$, by looking at subhypotheses, such as H_0: $\mu_1 = \mu_2$.

This involves multiple tests. However, the solution is not to use a simple t-test repeatedly for every possible combination taken two at a time. That, apart from introducing many tests, will considerably increase the significance level, the probability of type I error. For example, to test four samples we will need $\binom{4}{2} = 6$ tests. If each one of the comparisons is tested with the same value of $\alpha = P$ (type I error), and if all the null hypotheses involving six comparisons are true, then the probability of rejecting at least one of them is:

$$P(\text{at least one type I error}) = 1 - (1 - \alpha)^6.$$

In particular, if $\alpha = 0.01$, then P(at least one type I error) $= 0.077181$, which is significantly higher than the original specified error value of 0.01.

One way to investigate the problem is to use a multiple comparison procedure. A good deal of work has been done on problems of multiple comparisons. There are a variety of methods available in the literature, such as the Bonferroni procedure, Tukey's method, and Scheffe's method. We now describe one of the more popular procedures, called Tukey's method, for completely randomized, one-factor design.

In this multiple comparison problem, we would like to test H_0: $\mu_i = \mu_j$ versus H_a: $\mu_i \neq \mu_j$, for all $i \neq j$. Tukey's method will be used to test all possible differences of means to decide whether at least one of the differences $\mu_i - \mu_j$ is considerably different from zero. In this comparison problem, Tukey's method makes use of confidence intervals for $\mu_i - \mu_j$. If each confidence interval has a confidence level $1 - \alpha$, then the probability that all confidence intervals include their respective parameters is less than $1 - \alpha$. We now describe this method where each of the k sample means is based on the common number of observations, n.

Let $N = kn$ be the total number of observations and let

$$S^2 = \frac{1}{N-k} \sum_{i=1}^{k} \sum_{j=1}^{n_i=n} \left(Y_{ij} - \overline{T}_i\right)^2.$$

Let $\overline{T}_{\max} = \max\left(\overline{T}_1, \ldots, \overline{T}_k\right)$ and $\overline{T}_{\min} = \min\left(\overline{T}_1, \ldots, \overline{T}_k\right)$. Define the random variable

$$Q = \frac{\overline{T}_{\max} - \overline{T}_{\min}}{S\sqrt{n}}.$$

The distribution of Q under the null hypothesis H_0: $\mu_1 = \ldots = \mu_k$ is called the Studentized range distribution, which depends on the number of samples k and the degrees of freedom $\upsilon = N - k = (n-1)k$. We denote the upper α critical value by $q_{\alpha,k,\upsilon}$. The Studentized range distribution table gives values for selected values of k, υ, and α as 0.01, 0.05, and 0.9. The following theorem, attributable to Tukey, defines the test procedure.

Theorem 9.5.1 *Let* $\overline{T}_i$, $i = 1, 2, \ldots, k$, *be the* k *sample means in a completely randomized design. Let* μ_i, $i = 1, 2, \ldots, k$, *be the true means and let* $n_i = n$ *be the common sample size. Then the probability that all* $\binom{k}{2}$ *differences* $\mu_i - \mu_j$ *will simultaneously satisfy the inequalities*

$$\left(\overline{T}_i - \overline{T}_j\right) - q_{\alpha,k,\upsilon}\frac{s}{\sqrt{n}} \leq \mu_i - \mu_j \leq \left(\overline{T}_i - \overline{T}_j\right) + q_{\alpha,k,\upsilon}\frac{s}{\sqrt{n}},$$

is $(1 - \alpha)$, *where* $q_{\alpha,k,\upsilon}$ *is the upper* α *critical value of the Studentized range distribution. If, for a given* i *and* j, *zero is not contained in the preceding inequality,* H_0: $\mu_i = \mu_j$ *can be rejected in favor of* H_a: $\mu_i \neq \mu_j$, *at the significance level of* α.

Now we give a step-by-step approach to implementing Tukey's method discussed above.

Procedure to find $(1-\alpha)100\%$ confidence intervals for difference of means with common sample size N: Tukey's method

1. There are $\binom{k}{2}$ comparisons of μ_i versus μ_j.
2. Compute the following quantities:

$$\overline{T}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, i = 1, 2, \ldots, k,$$

and

$$s^2 = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i=n}\left(y_{ij}-\overline{T}_i\right)^2, \text{where } N = kn.$$

3. From the Studentized range distribution table, find the upper α critical value, $q_{\alpha,k,\upsilon}$, where $\upsilon = N - k = (n-1)k$.
4. For each $\binom{k}{2}$ pair (i, j), $i \neq j$, compute the Tukey interval:

$$\left(\left(\overline{T}_i - \overline{T}_j\right) - q_{\alpha,k,v}\frac{s}{\sqrt{n}}, \left(\overline{T}_i - \overline{T}_j\right) + q_{\alpha,k,v}\frac{s}{\sqrt{n}}\right).$$

5. Let NR denote insufficient evidence for rejecting H_0. Create the following table for each $\binom{k}{2}$ pairwise difference $\mu_i - \mu_j$, $i \neq j$, and *do not* reject if the Tukey interval contains the number 0. Otherwise reject.

Table 9.18 is used to summarize the final calculations of the Tukey method.

TABLE 9.18 Tuky Method-Calculation Summaries.

$\mu_i - \mu_j$	$\overline{T}_i - \overline{T}_j$	Tukey interval	Observation	Conclusion
$\mu_1 - \mu_2$	$\overline{T}_1 - \overline{T}_2$	…	Doesn't contain 0	Reject
$\mu_1 - \mu_3$	$\overline{T}_1 - \overline{T}_3$	…	Contains 0	Do not reject
⋮	⋮	⋮	⋮	⋮

In practice, there are now numerous statistical packages available for Tukey's purpose. The following example is solved using Minitab. The necessary Minitab commands are given in Example 9.7.3.

EXAMPLE 9.5.1

Table 9.19 shows the 1-year percentage total return of the top five stock funds for five different categories (source: *Money,* July 2000). Which categories have similar top returns and which are different? Use 95% Tukey's confidence intervals.

TABLE 9.19 Percentage Returns by Stock Types.

Large-cap	Mid-cap	Small-cap	Hybrid	Specialty
110.1	299.8	153.8	68.3	181.6
102.9	139.0	139.8	67.1	159.3
93.1	131.2	138.3	42.5	138.3
83.0	110.5	121.4	40.0	132.6
83.3	129.2	135.9	41.0	135.7

Solution

For simplicity of computation, we will use SPSS (Minitab steps are given in Example 9.7.2). The following is the output.

One-way
ANOVA
RETURN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	41,243.698	4	10,310.925	7.397	.001
Within Groups	27,877.580	20	1393.879		
Total	69,121.278	24			

Note that since the p *value is 0.001, we are rejecting the null hypothesis that all means are equal. To find out which of the means might be different, we use the multiple comparison output.*

Post Hoc Tests

Multiple Comparisons

Dependent Variable: RETURN
Tukey HSD

(I) FUND	(J) FUND	Mean Difference (I−J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
1.00	2.00	−67.4600	23.61253	.066	−138.1175	3.1975
	3.00	−43.3600	23.61253	.382	−114.0175	27.2975
	4.00	42.7000	23.61253	.396	−27.9575	113.3575
	5.00	−55.0200	23.61253	.177	−125.6775	15.6375
2.00	1.00	67.4600	23.61253	.066	−3.1975	138.1175
	3.00	24.1000	23.61253	.843	−46.5575	94.7575
	4.00	19.1600*	23.61253	.001	39.5025	180.8175
	5.00	12.4400	23.61253	.984	−58.2175	83.0975
3.00	1.00	43.3600	23.61253	.382	−27.2975	114.0175
	2.00	−24.1000	23.61253	.843	−94.7575	46.5575
	4.00	86.0600*	23.61253	.012	15.4025	156.7175
	5.00	−11.6600	23.61253	.987	−82.3175	58.9975
4.00	1.00	−42.7000	23.61253	.396	−113.3575	27.9575
	2.00	− 19.1600*	23.61253	.001	−180.8175	−39.5025
	3.00	−86.0600*	23.61253	.012	−156.7175	−15.4025
	5.00	− 97.7200*	23.61253	.004	−168.3775	−27.0625
5.00	1.00	55.0200	23.61253	.177	−15.6375	125.6775
	2.00	−12.4400	23.61253	.984	−83.0975	58.2175
	3.00	11.6600	23.61253	.987	−58.9975	82.3175
	4.00	97.7200*	23.61253	.004	27.0625	168.3775

*The mean difference is significant at the 0.05 level.

Homogeneous Subsets

RETURN

Tukey HSD[a]

FUND	N	Subset for alpha = .05: 1	2
4.00	5	51.7800	
1.00	5	94.4800	94.4800
3.00	5		137.8400
5.00	5		149.5000
2.00	5		161.9400
Sig.		.396	.066

Means for groups in homogeneous subsets are displayed.
[a] Uses Harmonic Mean Sample Size = 5.000.

The Tukey intervals for pairwise differences ($\mu_i - \mu_j$) are in the foregoing computer printout. For example, the Tukey interval for ($\mu_1 - \mu_2$) is (−138.1, 3.2) and for ($\mu_2 - \mu_4$) is (39.5, 180.8). Also, sample mean and standard deviation are given in the output. For example, 94.48 is the sample mean of the five data points of large-cap funds, and 11.97 is the sample standard deviation of the five data points of large-cap funds.

If the Tukey interval for a particular difference ($\mu_j - \mu_i$) contains the number 0, we do not reject H_0: $\mu_i = \mu_j$. Otherwise, we reject H_0: $\mu_i = \mu_j$. For example, the interval for ($\mu_4 - \mu_2$) is (39.5−180.8) and does not contain 0. Hence, we reject H_0: $\mu_4 = \mu_2$.

The complete table corresponding to step 5 is produced in Table 9.20, where NR represents "do not reject."

TABLE 9.20 Tukey Intervals and Decisions.

$\mu i - \mu j$	$\overline{T}_i - \overline{T}_j$	Tukey interval	R or NR	Conclusion
$\mu_1 - \mu_2$	161.94 − 94.48	(−138.1, 3.2)	*NR*	$\mu_1 = \mu_2$
$\mu_1 - \mu_3$	137.84 − 94.48	(−114.0, 27.3)	*NR*	$\mu_1 = \mu_3$
$\mu_2 - \mu_3$	137.84−161.94	(−46.6, 94.8)	*NR*	$\mu_3 = \mu_2$
$\mu_1 - \mu_4$	51.78−94.48	(−27.9, 113.3)	*NR*	$\mu_4 = \mu_1$
$\mu_2 - \mu_4$	51.78−161.94	(39.5, 180.8)	*R*	$\mu_4 \neq \mu_1$
$\mu_3 - \mu_4$	51.78−137.84	(15.4, 156.7)	*R*	$\mu_4 \neq \mu_3$
$\mu_1 - \mu_5$	149.50−94.98	(−125.6, 15.6)	*NR*	$\mu_5 = \mu_1$
$\mu_2 - \mu_5$	149.50−161.94	(−58.2, 83.1)	*NR*	$\mu_5 = \mu_2$
$\mu_3 - \mu_5$	149.50−137.84	(−82.3, 59.0)	*NR*	$\mu_5 = \mu_3$
$\mu_4 - \mu_5$	149.50−51.78	(−168.3,−27.1)	*R*	$\mu_5 \neq \mu_4$

NR, do not reject; *R*, Reject.

Based on the 95% Tukey intervals, the average top return of hybrid funds is different from those for mid-cap, small-cap, and specialty funds. All other returns are similar.

In Tukey's method, the confidence coefficient for the set of all pairwise comparisons $\{\mu_i - \mu_j\}$ is exactly equal to $1 - \alpha$ when all sample sizes are equal. For unequal sample sizes, the confidence coefficient is greater than $1 - \alpha$. In this sense, Tukey's procedure is conservative when the sample sizes are not equal. In the case of unequal sample sizes, one has to estimate the standard deviation for each pairwise comparison. Tukey's procedure for unequal sample sizes is sometimes referred to as the *Tukey−Kramer method*.

Exercises 9.5

9.5.1 A large insurance company wants to determine whether there is a difference in the average time to process claim forms among its four different processing facilities. The data in Table 9.21 represent weekly average number of days to process a form over a period of 4 weeks.

(a) Test whether there is a difference in the average processing times at the 0.05 level.
(b) Test whether there is a difference, using Tukey's method to find which facilities are different.
(c) Interpret your results and state any assumptions you have made in solving the problem.

TABLE 9.21 Claim Processing Time by Facility

Facility 1	Facility 2	Facility 3	Facility 4
1.50	2.25	1.30	2.0
0.9	1.85	2.75	1.5
1.12	1.45	2.15	2.85
1.95	2.15	1.55	1.15

9.5.2 Table 9.22 gives the rental vacancy rates by US region (source: US Census Bureau) for 5 years.

(a) Test at the 0.01 level whether the true rental vacancy rates by region are the same for all 5 years.

(b) If there is a difference, use Tukey's method to find which regions are different.

TABLE 9.22 Retal Vacancy by Year for Different Regions

Rental units	1995	1996	1997	1998	1999
Northeast	7.2	7.4	6.7	6.7	6.3
Midwest	7.2	7.9	8.0	7.9	8.6
South	8.3	8.6	9.1	9.6	10.3
West	7.5	7.2	6.6	6.7	6.2

9.5.3 Table 9.23 gives lower limits of income (approximated to nearest $1000 and calculated as of March of the following year) by race for the top 5% of US households from 1994 to 1998 (source: US Census Bureau).

(a) Test at the 0.05 level whether the true lower limits of income for the top 5% of US households for each race are the same for all 5 years.

(b) If there is a difference, use Tukey's method to find which is different.

(c) Interpret your results and state any assumptions you have made in solving the problem.

TABLE 9.23 Income Lower Limits for Top 5% by Race

Race	1994	1995	1996	1997	1998
All races	110	113	120	127	132
White	113	117	123	130	136
Black	81	80	85	87	94
Hispanic	82	80	86	93	98

9.5.4 The data in Table 9.24 represent the mean serum cholesterol levels (given in milligrams per deciliter) by race and age in the United States from 1978 to 1980 (source: "Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults," *Arch. Intern. Med.* 148, January 1988).

TABLE 9.24 Mean Serum Cholostorel Level by Age and Race

	Age					
Race	**20–24**	**25–34**	**35–44**	**45–54**	**55–64**	**65–74**
All races	180	199	217	227	229	221
White	180	199	217	227	230	222
Black	171	199	218	229	223	217

(a) Test at the 0.01 level whether the true mean cholesterol levels for all races in the United States during 1978–80 are the same.

(b) If there is a difference, use Tukey's method to find which of the races are different with respect to mean cholesterol levels.

9.6 Chapter summary

In this chapter, we have introduced the basic idea of analyzing data from various experimental designs. In Section 9.3, we explained the one-way ANOVA for the hypothesis testing problem for more than two means (different treatments being applied, or different populations being sampled). The two-way ANOVA, having b blocks and k treatments consisting of b blocks of k experimental units each, is discussed in Section 9.4. We also described one popular procedure called Tukey's method for completely randomized, one-factor design for multiple comparisons. We saw in Chapter 8 that there are other possible designs, such as the Latin square design or Taguchi methods. We refer to specialized books on experimental design (Hicks and Turner) for more details on how to conduct ANOVA on such designs. In the final section, we give some computational examples.

We now list some of the key definitions introduced in this chapter:

- Completely randomized experimental design
- Randomized block design
- Studentized range distribution
- Tukey−Kramer method

In this chapter, we also learned the following important concepts and procedures:

- ANOVA procedure for two treatments
- One-way ANOVA procedure for $k \geq 2$ populations
- Procedure to find $(1-\alpha)100\%$ confidence intervals for difference of means with common sample size n; Tukey's method
- Computational procedure for randomized block design

9.7 Computer examples

Minitab, SPSS, SAS, and other statistical programming packages are especially useful when we perform an ANOVA. As we have experienced in earlier sections, an ANOVA computation is very tedious to complete by hand.

9.7.1 Examples using R

EXAMPLE 9.7.1 One-way ANOVA

The three random samples in the following table are independently obtained from three different normal populations with equal variances. At the $\alpha = 0.05$ level of significance, test for equality of means.

Sample	64	84	75	77	80	56	74	69	81	92	84
Group	1	1	1	1	1	2	2	2	3	3	3

This example assumes you have stored the data into two variables, sample and group. Please modify your code appropriately.

R-code

```
model = lm(sample~as.factor(group));
anova(model);
```

Notice we must use as.factor() to get the proper degrees of freedom.

Output

```
Analysis of Variance Table
Response: sample
                  Df  Sum     Sq Mean   Sq F value  Pr(>F)
as.factor(group)  2   560.67  280.333   4.8403      0.04192 *
Residuals         8   463.33   57.917
- - -
Signif. codes:    0   '***'   0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p value is less than 0.05, we reject H_0 of equal means.

EXAMPLE 9.7.2 Two-way ANOVA

A furniture company wants to know whether there are differences in stain resistance among the four chemicals used to treat three different fabrics. The following table shows the yields on resistance to stain (a low value indicates good stain resistance). At the $\alpha = 0.05$ level of significance, is there evidence to conclude that there is a difference in mean resistance among the four chemicals? Is there any difference in the mean resistance among the materials?

Chemical	1	2	3	4	1	2	3	4	1	2	3	4
Resistance	3	9	2	7	7	11	5	9	6	8	7	8
Material	1	1	1	1	2	2	2	2	3	3	3	3

This example assumes you have stored data into three variables chemical, resistance, and material. Please modify your code appropriately.

R-code

```
model = lm(resistance~as.factor(chemical)+as.factor(material));
anova(model);
```

Output

```
Analysis of Variance Table
Response:  resistance
                      Df  Sum Sq  Mean Sq  F value  Pr(>F)
as.factor(chemical)    3  43.667  14.5556   7.3803  0.01943 *
as.factor(material)    2  16.167   8.0833   4.0986  0.07548.
Residuals              6  11.833   1.9722
—
Signif. codes:   0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

p values suggest that the chemical is a significant factor but the material is not.

EXAMPLE 9.7.3 Tukey's Method

The following table shows the 1-year percentage total return of the top five stock funds for five different categories (source: *Money*, July 2000). Which categories have similar top returns and which are different? Use 95% Tukey's confidence intervals. This example assumes you have stored your data into "stocks" and "groups" variables that pair.

Large-cap	Mid-cap	Small-cap	Hybrid	Specialty
19.1	299.8	153.8	68.3	181.6
102.9	139.0	139.8	67.1	159.3
93.1	131.2	138.3	42.5	138.3
83.0	19.5	121.4	40.0	132.6
83.3	129.2	135.9	41.0	135.7

This assumes your "stocks" variable is typed in column by column, top to bottom.

R-code

```
groups=c(rep("Large-cap",5),rep("Mid-cap",5),rep("Small-
cap",5),rep("Hybrid",5),rep("Specialty",5));
model=aov(stocks~as.factor(groups));
TukeyHSD(model);
```

Output

```
Tukey multiple comparisons of means
95% family-wise confidence level
Fit: aov(formula = stocks ~ as.factor(groups))
```

$'as.factor(groups)'

	diff	lwr	upr	p adj
Large-cap-Hybrid	42.70	−27.957536	113.35754	0.3963504
Mid-cap-Hybrid	19.16	39.502464	180.81754	0.0012546*
Small-cap-Hybrid	86.06	15.402464	156.71754	0.0124242*
Specialty-Hybrid	97.72	27.062464	168.37754	0.0041271*
Mid-cap-Large-cap	67.46	−3.197536	138.11754	0.0657451
Small-cap-Large-cap	43.36	−27.297536	114.01754	0.3816028
Specialty-Large-cap	55.02	−15.637536	125.67754	0.1765264
Small-cap-Mid-cap	−24.10	−94.757536	46.55754	0.8429013
Specialty-Mid-cap	−12.44	−83.097536	58.21754	0.9835150
Specialty-Small-cap	11.66	−58.997536	82.31754	0.9870429

This shows that mean returns for hybrid to mid-cap, small-cap, and specialty are different.

9.7.2 Minitab examples

EXAMPLE 9.7.4

(One-way ANOVA): The three random samples in Table 9.25 are independently obtained from three different normal populations with equal variances.

TABLE 9.25 Three Random Samples from Different Normal Populations.

Sample 1	Sample 2	Sample 3
64	56	81
84	74	92
75	69	84
77		
80		

At the $\alpha = 0.05$ level of significance, test for equality of means.

Solution

Enter sample 1 data in C1, sample 2 in C2, and sample 3 in C3.

Stat > ***ANOVA*** > ***One-way (unstacked)*** … > in ***Responses (in separate columns)***: type ***C1 C2 C3*** and click ***OK***.

We get the following output:

One-Way Analysis of Variance

```
Analysis of Variance

Source      DF        SS        MS        F        P
Factor       2     560.7     280.3     4.84    0.042
Error        8     463.3      57.9
Total       10    1024.0

                              Individual 95% CIs For Mean
                              Based on Pooled StDev

Level          N    Mean   StDev  --+---------+---------+---------+-
C1             5  76.000   7.517           (----*-----)
C2             3  66.333   9.292     (----*-----)
C3             3  85.667   5.686     (----*-----)
                                --+---------+---------+---------+--
Pooled StDev = 7.610   60      72        84        96
```

We can see that the output contains SS, MS, individual column means, and standard deviation values. Also, the F *value gives the value of the test statistic, and the* p *value is obtained as 0.042. Comparing this* p *value of 0.042 with* $\alpha = 0.05$, *we will reject the null hypothesis.*

If we want to create side-by-side box plots to graphically test homogeneity of variances, we can do the following.

Enter all the data (from all three samples) in ***C1***, *and enter the sample identifier number in* ***C2*** *(that is, 1 if the data belong to sample 1, 2 for sample 2, and 3 for sample 3).*

Graph > ***Boxplot*** > in ***Y*** column, type ***C1*** and in ***X*** column, type ***C2*** > click ***OK***.

Then as in Example 9.3.2, interpret the resulting box plots.

EXAMPLE 9.7.5

Give Minitab steps for randomized block design for the data of Example 9.4.1.

Solution

To put the data into the format for Minitab, place all the data values in one column (say, ***C2***). *Let numbers 1, 2, 3, 4 represent the chemicals and numbers 1, 2, 3 represent the fabric material. In one column (say,* ***C1***) *place numbers 1 through 4 with respect to the data values identifying the factor (chemical) used. In another column (say,* ***C3***) *place corresponding numbers 1 through 3 to identify the second factor (material) used. See Table 9.26.*

TABLE 9.26 Example of Randomized Block Design.

C1 chemical	C2 response	C3 material
1	3	1
2	9	1
3	2	1
4	7	1
1	7	2
2	11	2
3	5	2
4	9	2
1	6	3
2	8	3
3	7	3
4	8	3

Then do the following:

Stat > ***ANOVA*** > ***Two-way*** ... > in ***Response:*** type ***C2***, in ***Row Factor***: type ***C1***, and in ***Column factor***: type ***C3*** > ***OK***

We will get the following output.

Two-Way Analysis of Variance

```
Analysis of variance for Response
```

Source	DF	SS	MS	F	P
Chemical	3	43.67	14.56	7.38	0.019
Material	2	16.17	8.08	4.10	
Error	6	11.83	1.97		
Total	11	71.67			

Note that the output contains p *values for the effects both of the chemicals and of the materials. Because the* p *value of 0.019 is less than* $\alpha = 0.05$, *we reject the null hypothesis and conclude that there is a difference in mean resistance among the four chemicals. For the materials, the* p *value of 0.075 is greater than* $\alpha = 0.05$, *so we cannot reject the null hypothesis and conclude that there is no difference in the mean resistance among the materials.*

EXAMPLE 9.7.6

Give the Minitab steps for using Tukey's method for the data of Example 9.5.1.

Solution

To use Tukey's method, it is necessary to enter the data in a particular way. Enter all the data points in column ***C1****: the first five from large-cap, next five from mid-cap, and so on, with the last five from specialty. In column* ***C2****, enter the number identifying the data points: the first four numbers are 1 (identifying 1 as the data belonging to large-cap), next five numbers are 2, and so on; the last five numbers are 5. Then:*

Stat > ***ANOVA*** > ***One-way*** … > ***Comparisons*** … > click ***Tukey's, family error rate****:* and type ***5*** (to represent 100α% error) > ***OK*** > in ***Response****:* type ***C1***, and in ***Factor****:* type ***C2*** > ***OK***

We will get an output similar to that given in the solution part of Example 9.5.1. *For discussion of the output, refer to Example 9.5.1*.

9.7.3 SPSS examples

EXAMPLE 9.7.7

Conduct a one-way ANOVA for the data of Example 9.7.1. Use $\alpha = 0.05$ level of significance, and test for equality of means.

Solution

In SPSS, we need to enter the data in a special way. First name column ***C1*** *as Sample and column* ***C2*** *as Values. In the Sample column, enter the numbers to identify from which group the data come. In this case, enter 1 in the first five rows, 2 in the next three rows, and 3 in the last three rows. In the Values column, enter sample 1 data in the first five rows, sample 2 data in the next five rows, and sample 3 data in the last three rows. Then:*

Analyze > ***Compare Means*** > ***One-way ANOVA*** … > Bring ***Values*** to ***Dependent List****:* and ***Sample*** to ***Factor****:* > ***OK***

EXAMPLE 9.7.8

Give the SPSS steps for using Tukey's method for the data of Example 9.5.1.

Solution

First name column C1 as ***Fund*** *and column C2 as* ***Return****. In the Fund column, enter the numbers to identify from which group the data come. In this case, the first four numbers are 1 (identifying 1 as the data belonging to large-cap), the next four numbers are 2, and so on, until the last four numbers are 5. In the Return column, enter large-cap return data in the first four rows, mid-cap data in the next four rows, and so on, the last four from specialty. Then:*

Analyze > ***Compare Means*** > ***One-way ANOVA*** … > Bring ***Return*** to ***Dependent List****:* and ***Fund*** to ***Factor****:* > Click ***Post-Hoc*** … > click ***Tukey*** > click ***Continue*** > *OK*

We will get the output as in Example 9.5.1.

Interpretation of the output is given in Example 9.5.1. When the treatment effects are significant, as in this example where the p *value is 0.001, the means must then be further examined to determine the nature of the effects. There are procedures called post hoc tests to assist the researcher in this task. For example, looking at the output column Sig., we could observe that there are significant differences in the mean returns between funds 2 and 4 and funds 4 and 5.*

9.7.4 SAS examples

EXAMPLE 9.7.9

Using SAS, conduct a one-way ANOVA for the data of Example 9.7.1. Use $\alpha = 0.05$ level of significance, and test for equality of means.

Solution

We could use the following code.

```
Options nodate nonumber;
options ls=80 ps=50;
DATA Scores;
INPUT Sample Value @@;
DATALINES;
```

1	64	1	84	1	75	1	77	1	80
2	56	2	74	2	69				
3	81	3	92	3	84				

```
;
PROC ANOVA DATA=Scores;
TITLE 'ANOVA for Scores';
CLASS Sample;
MODEL Value = Sample;
MEANS Sample;
RUN;
```

We could have used PROC GLM instead of PROC ANOVA to perform the ANOVA procedure. Usually, PROC ANOVA is used when the sizes of the samples are equal; otherwise PROC GLM is more desirable. The next example will show how to do the multiple comparison using Tukey's procedure.

EXAMPLE 9.7.10

Give the SAS commands for using Tukey's method for the data of Example 9.5.1.

Solution

We could use the following code.

```
Options nodate nonumber;
options ls=80 ps=50;
DATA Mfundrtn;
INPUT Fund Return @@;
DATALINES;
```

1 19.1	2 299.8	3 153.8	4 68.3	5 181.6
1 102.9	2 139.0	3 139.8	4 67.1	5 159.3
1 93.1	2 131.2	3 138.3	4 42.5	5 138.3
1 83.3	2 129.2	3 135.9	4 41.0	5 135.7
1 83.0	2 19.5	3 121.4	4 40.0	5 132.6

```
;
PROC GLM DATA=Mfundrtn;
TITLE 'ANOVA for Mutual fund returns';
CLASS Fund;
MODEL Return=Fund;
MEANS Fund / tukey;
RUN;
```

```
                 ANOVA for Mutual fund returns

                       The GLM Procedure

            Class Level Information

           Class    Levels    Values

           Fund      5       1 2 3 4 5

           Number of observations   25

           ANOVA for Mutual fund returns
                 The GLM Procedure

Dependent Variable: Return
                     Sum of
Source         DF   Squares   Mean Square  F Value  Pr > F

Model           4  41243.69840  1039.92460   7.40    0.0008

Error          20 27877.58000  1393.87900

Corrected Total 24 69121.27840

          R-Square    Coeff Var    Root MSE    Return Mean

          0.596686     31.34524     37.33469    119.1080

Source     DF   Type I SS   Mean Square  F Value   Pr > F

Fund        4   41243.69840 1039.92460    7.40    0.0008

Source     DF   Type III SS Mean Square  F Value   Pr > F

Fund        4   41243.69840 1039.92460    7.40    0.0008

                 ANOVA for Mutual fund returns

                       The GLM Procedure

         Tukey's Studentized Range (HSD) Test for Return
NOTE: This test controls the Type I experiment wise error rate,
but it generally has a higher Type II error rate than REGWQ.
          Alpha                                  0.05
          Error Degrees of Freedom                 20
          Error Mean Square                  1393.879
          Critical Value of Studentized Range 4.23186
          Minimum Significant Difference       70.658
```

```
     Means with the same letter are not significantly different.

        Tukey Grouping         Mean      N    Fund

                    A       161.94       5     2
                    A
                    A       149.50       5     5
                    A
                    A       137.84       5     3
                    A

              B     A        94.48       5     1
              B
              B              51.78       5     4

                          The GLM Procedure

            Tukey's Studentized Range (HSD) Test for Value

NOTE: This test controls the Type I experiment wise error rate,
but it generally has a higher Type II error rate than REGWQ.

          Alpha                                    0.05
          Error Degrees of Freedom           20
          Error Mean Square                    1393.879
          Critical Value of Studentized Range 4.23186
          Minimum Significant Difference         70.658

          Means with the same letter are not significantly
          different.

     Turkey Grouping      Mean      N    Sample

                   A     161.94    5       2
                   A
                   A     149.50    5       5
                   A
                   A     137.84    5       3
                   A
               B   A     94.48     5       1
               B
               B         51.78     5       4
```

Looking at the p *value of 0.008, which is less than* $\alpha = 0.05$*, we conclude that there is a difference in mutual fund returns.*

In the previous example, we used the post hoc test Tukey. We could have used other options such as DUNCAN, SNK, LSD, and SCHEFFE. The test is performed at the default value of $\alpha = 0.05$*. If we want to specify, say,* $\alpha = 0.01$*, or* 0.1*, we could have done so by using the command MEANS Fund / Tuckey ALPHA=0.01.*

If we need all the confidence intervals in the Tukey method, in the code just given, we have to modify 'MEANS Fund / Tukey;' to 'MEANS Fund / LSD TUKEY CLDIFF;' which will result in the following output.

```
          ANOVA for Mutual fund returns
                The GLM Procedure
             Class Level Information
         Class      levels       Values
         Fund          5    1 2 3 4 5
            Number of observations 25
                    ANOVA for Mutual fund returns
                          The GLM Procedure
                     Dependent Variable: Return
                 Sum of
Source    DF   Squares      Mean Square    F Value  Pr > F
Model      4   41243.69840    1039.92460      7.40  0.0008
Error     20   27877.58000      1393.87900
Corrected Total 24  69121.27840
     R-Square         Coeff Var    Root MSE  Return Mean
     0.596686         31.34524      37.33469   119.1080
Source    DF   Type I SS     Mean Square  F Value  Pr > F
Fund       4   41243.69840   1039.92460    7.40    0.0008
Source    DF   Type III SS   Mean Square  F Value  Pr > F
Fund       4   41243.69840   1039.92460    7.40    0.0008
                    ANOVA for Mutual fund returns
                          The GLM Procedure
                       t-tests (LSD) for Return
```

```
NOTE: This test controls the Type I comparison wise error rate,
not the experiment wise error rate.

     Alpha                                  0.05
     Error Degrees of Freedom                 20
     Error Mean Square                  1393.879
     Critical Value of t                 2.08596
     Least Significant Difference         49.255

Comparisons significant at the 0.05 level are indicated by ***.

                           Difference

       Fund    Between        95% Confidence
    Comparison   Means            Limits

    2 - 5        12.44       -36.81     61.69
    2 - 3        24.10       -25.15     73.35
    2 - 1        67.46        18.21    116.71   ***
    2 - 4        19.16        60.91    159.41   ***
    5 - 2       -12.44       -61.69     36.81
    5 - 3        11.66       -37.59     60.91
    5 - 1        55.02         5.77    104.27   ***
    5 - 4        97.72        48.47    146.97   ***
    3 - 2       -24.10       -73.35     25.15
    3 - 5       -11.66       -60.91     37.59
    3 - 1        43.36        -5.89     92.61
    3 - 4        86.06        36.81    135.31   ***
    1 - 2       -67.46      -116.71    -18.21   ***
    1 - 5       -55.02      -104.27     -5.77   ***
    1 - 3       -43.36       -92.61      5.89
    1 - 4        42.70        -6.55     91.95
    4 - 2       -19.16      -159.41    -60.91   ***
    4 - 5       -97.72      -146.97    -48.47   ***
    4 - 3       -86.06      -135.31    -36.81   ***
    4 - 1       -42.70       -91.95      6.55

             ANOVA for Mutual fund returns

                   The GLM Procedure

   Tukey's Studentized Range (HSD) Test for Return

   NOTE: This test controls the Type I experiment wise
error rate.

     Alpha                                       0.05
     Error Degrees of Freedom                      20
     Error Mean Square                       1393.879
     Critical Value of Studentized Range      4.23186
     Least Significant Difference            70.658
```

Comparisons significant at the 0.05 level are indicated by ***.

Fund Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
2 - 5	12.44	-58.22	83.10	
2 - 3	24.10	-46.56	94.76	
2 - 1	67.46	-3.20	138.12	
2 - 4	19.16	39.50	180.82	***
5 - 2	-12.44	-83.10	58.22	
5 - 3	11.66	-59.00	82.32	
5 - 1	55.02	-15.64	125.68	
5 - 4	97.72	27.06	168.38	***
3 - 2	-24.10	-94.76	46.56	
3 - 5	-11.66	-82.32	59.00	
3 - 1	43.36	-27.30	114.02	
3 - 4	86.06	15.40	156.72	***
1 - 2	-67.46	-138.12	3.20	
1 - 5	-55.02	-125.68	15.64	
1 - 3	-43.36	-114.02	27.30	
1 - 4	42.70	-27.96	113.36	
4 - 2	-19.16	-180.82	-39.50	***
4 - 5	-97.72	-168.38	-27.06	***
4 - 3	-86.06	-156.72	-15.40	***
4 - 1	-42.70	-113.36	27.96	

Exercises 9.7

9.7.1. Using the data of Exercise 9.5.4, perform a one-way ANOVA using any of the softwares (R, Minitab, SPSS, or SAS).

9.7.2. Using the data of Exercise 9.5.2, perform Tukey's test using any of the softwares (R, Minitab, SPSS, or SAS).

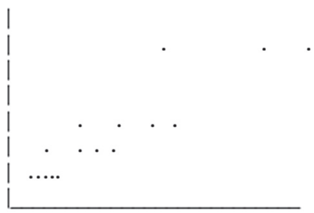
9.7.3. Using the data of Exercise 9.5.4, perform Tukey's test using any of the softwares (R, Minitab, SPSS, or SAS).

Projects for Chapter 9

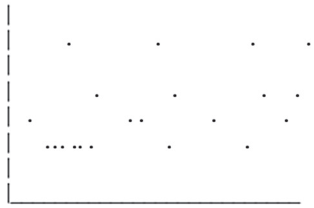
9A Transformations

The basic model for the ANOVA requires that the independent observations come from normal populations with equal variances. These requirements are rarely met in practice, and the extent to which they are violated affects the validity of the subsequent inference. Therefore, it is important for the investigator to decide whether the assumptions are at least approximately satisfied and, if not, what can be done to rectify the situation. Hence, it is necessary to (1) examine the data for marked departures from the model and, if necessary, (2) apply an appropriate transformation to the data to bring them more in line with the basic assumptions.

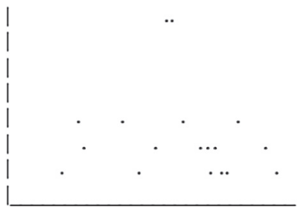
A simple way to check for the equality of the population variances is to calculate the sample variances and plot against mean as in Fig. 9.3. If the graph suggests a relation between sample mean and variance, then the relation very likely exists between population mean and variance, and hence, the population from which the samples are taken may very well be nonnormal and/or the data are heterogeneous. A simple visual check of heterogeneity can be done using the following type of scatterplot of mean versus variance across replicates.



Heterogeneous variance where the variance is proportional to the mean.



Homogeneous variance



Outliers – transformation will not work.

If a study of sample means and variances reveals a marked departure from the model, the observations may be transformed into a new set to which the methods of ANOVA are better suited. Three commonly used transformations are the following:

(a) ***The logarithmic transformation***: This is used if the graph of sample means against sample variance suggests a relation of the form:

$$s^2 = C\left(\overline{X}^2\right),$$

That is, if $\sigma^2 = k\mu^2$, replace each observation X with its logarithm to the base 10,

$$Y = log_{10}X;$$

or, if some X values are 0, with $Y = \log_{10}(X + 1)$.

(b) ***The square root transformation:*** This is used if the relation is of the form:

$$s^2 = C\overline{X}$$

That is, if $\sigma^2 = k\mu$, replace X with its square root,

$$Y = \sqrt{X}$$

or, if the values of X are very close to 0, with the square root of (X + ½). This relation is found in data from Poisson populations, where the variance is equal to the mean.

(c) ***The angular transformation:*** If the observations are counts of a binomial nature, and $\widehat{p}$ is the observed proportion, replace $\widehat{p}$ with:

$$\theta = \arcsin\sqrt{\widehat{p}},$$

which is the principal angle (in degrees or radians) whose sine is the square root of $\widehat{p}$.

(i) To check for the equality of the population variances, calculate the sample variances for each of the data sets given in the exercises of Section 9.3 and plot against the corresponding mean.
(ii) If there is assumptional violation, perform one of the transformations described earlier and do the ANOVA procedure for the transformed data.

9B Analysis of variance with missing observations

In the two-way ANOVA, we assumed that each block cell has one treatment value. However, it is possible that some observations in some block cells may be missing for various reasons, such as if the investigator failed to record the observations, the subject discontinued participation in the experiment, or the subject moved to a different place or died prior to completion of the experiment. In those cases, this project gives a method of inserting estimates of the missing values.

Let $y_{..}$ denote the total of all kb observations. If the observation corresponding to the ith row and the jth column, which is denoted by $y_{ij\cdot}$, is missing, then all the sums of squares are calculated as before, except that the y_{ij} term is replaced by:

$$\widehat{y}_{ij} = \frac{bB'_j + kT'_i - y'_{..}}{(k-1)(b-1)},$$

where T'_i denotes the total of $b-1$ observations in the ith row, B'_j denotes the total of $k-1$ observations in the jth column, and $y'_{..}$ denotes the sum of all $kb-1$ observations. Using calculus, one can show that $\widehat{y}_{ij}$ minimizes the error sum of squares. One should not include these estimates when computing relevant degrees of freedom. With these changes, proceed to perform the analysis as in Section 9.4. For more details on the method, refer to Sahai and Ageel (2000), p. 145.

Perform the test of Example 9.4.1, now with a missing value for material III and chemical C_4. Does the conclusion change?

9C Analysis of variance in linear models

To determine whether the multiple regression model introduced in Section 8.5 is adequate for predicting values of dependent variable y, one can use the ANOVA F-test. The model is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon,$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \sim N(0, \sigma^2)$ and ε_i and ε_j are uncorrected if $i \neq j$. Define the multiple coefficient of determination, R^2, as:

$$R^2 = 1 - \frac{\sum (y_i - \widehat{y}_i)^2}{\sum (y_i - \overline{y})^2}.$$

The ANOVA F-test:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0 \text{ versus}$$

$$H_a: \text{At least one of the parameters, } \beta_1, \beta_2, \ldots, \beta_k, \text{differs from 0.}$$

Test statistic:

$$F = \frac{\text{Mean square for model}}{\text{Mean square for error}}$$

$$= \frac{SS(\text{model})/k}{SSE/[n-(k+1)]}$$

$$= \frac{R^2/k}{(1-R^2)/[n-(k+1)]},$$

where

$$n = \text{number of observations}$$

$$k = \text{number of parameters in the model excluding } \beta_0.$$

From the F-table, determine the value of F_α with k numerator degrees of freedom and $n - (k+1)$ denominator degrees of freedom. Then the rejection region is $\{F > F_\alpha\}$.

If we reject the null hypothesis, then the model can be taken as useful in predicting values of y.

Using the data of Example 8.5.1, test the overall utility of the fitted model:

$$y = 66.12 - 0.3794X_1 + 21.4365X_2$$

using the F-test described earlier.