

8-5 FURTHER ANALYSIS:

You have rejected the ANOVA null hypothesis. What next? This is an important question often overlooked in elementary introductions to analysis of variance. After all, what is the meaning of the statement “not all r population means are equal” if we cannot tell *in what way* the population means are not equal? We need to know which of our population means are large, which of them are small, and the magnitudes of the differences among them. These issues are addressed in this section.

ANOVA can be viewed as a machine or a box: in go the data, and out comes a conclusion—“all r population means are equal” or “not all r population means are equal.” If the ANOVA null hypothesis, $H_0: \mu_1 = \mu_2 = \dots = \mu_r$, is accepted and we therefore state that there is no strong evidence to conclude that differences exist

⁸ R. E. Krapfel, “An Advocacy Behavior Model of Organizational Buyers’ Vendor Choice,” *Journal of Marketing* (Fall 1985).

among the r population means, there is nothing more to say or do (unless, of course, you believe that differences do exist and you are able to gather more information to prove so). If the ANOVA null hypothesis is rejected, then we have evidence that the r population means are not all equal. This calls for *further analysis*—other hypothesis tests and/or the construction of confidence intervals to determine where the differences exist, their directions, and their magnitudes. The schematic diagram of the “ANOVA box” is shown in Figure 8–12.

Several methods have been developed for further analysis following the rejection of the null hypothesis in ANOVA. All the methods make use of the following two properties.

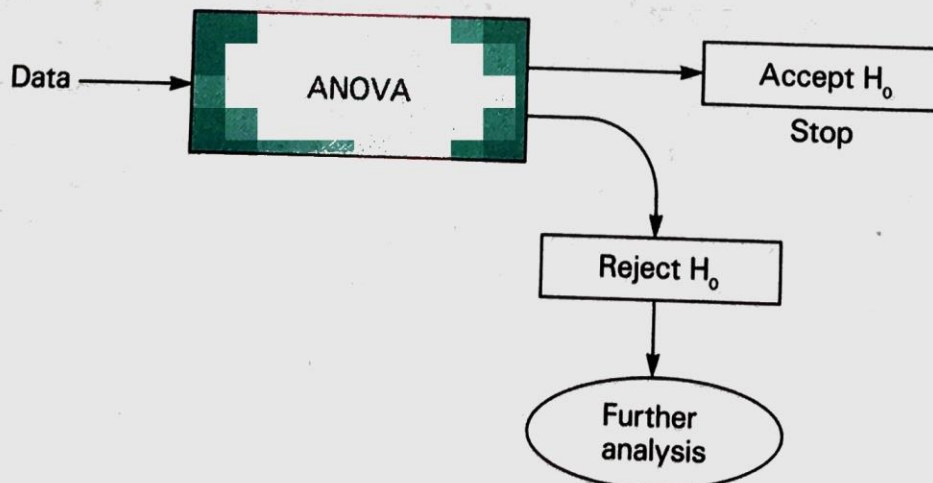
The sample means, $\bar{X}_i$, are unbiased estimators of the corresponding population means, μ_i .

The mean square error, MSE, is an unbiased estimator of the common population variance, σ^2 .

Since MSE can be read directly from the ANOVA table, we have another advantage of using an ANOVA table. This extends the usefulness of the table beyond the primary stages of analysis. The first and simplest post-ANOVA analysis is the estimation of separate population means, μ_i . It can be shown that, under the assumptions of analysis of variance, each sample mean, $\bar{X}_i$, has a normal distribution with mean μ_i and standard deviation $\sigma/\sqrt{n}$, where σ is the common standard deviation of the r populations. Since σ is not known, we estimate it by $\sqrt{\text{MSE}}$. We get the following relation.

$$\frac{\bar{X}_i - \mu_i}{\sqrt{\text{MSE}/\sqrt{n_i}}} \text{ has a } t \text{ distribution with } n - r \text{ degrees of freedom} \quad (8-19)$$

FIGURE 8–12 The ANOVA Diagram



This property leads us to the possibility of constructing confidence intervals for individual population means.

A $(1 - \alpha)$ 100% confidence interval for μ_i , the mean of population i :

$$\bar{x}_i \pm t_{\alpha/2} \frac{\sqrt{\text{MSE}}}{\sqrt{n_i}} \quad (8-20)$$

where $t_{\alpha/2}$ is the value of the t distribution with $n - r$ degrees of freedom that cuts off a right-hand-tail area equal to $\alpha/2$

Confidence intervals given by Equation 8-20 are included in the standard output of ANOVA in many packages, such as MINITAB. This will be seen in Section 8-9.

We now demonstrate the use of Equation 8-20 with the continuation of Example (b), the Club Med example. From Table 8-4, we get the sample means, $\bar{x}_i$:

Guadeloupe: $\bar{x}_1 = 89$

Martinique: $\bar{x}_2 = 75$

Haiti: $\bar{x}_3 = 73$

Paradise Is.: $\bar{x}_4 = 91$

St. Lucia: $\bar{x}_5 = 85$

From Table 8-6, the ANOVA table for this example, we get: $\text{MSE} = 504.4$ and degrees of freedom for error = $n - r = 195$. We also know that the sample size in each group is $n_i = 40$ for all $i = 1, \dots, 5$. Since a t distribution with 195 degrees of freedom is, for all practical purposes, a standard normal distribution, we use $z = 1.96$ in constructing 95% confidence intervals for the population mean responses of vacationers on the five islands. We will construct a 95% confidence interval for the mean response on Guadeloupe and leave the construction of the other four confidence intervals as an exercise. For Guadeloupe, we have the following 95% confidence interval for the population mean (μ_1):

$$\bar{x}_1 \pm t_{\alpha/2} \frac{\sqrt{\text{MSE}}}{\sqrt{n_1}} = 89 \pm 1.96 \frac{\sqrt{504.4}}{\sqrt{40}} = [82.04, 95.96]$$

The real usefulness of ANOVA, however, does not lie in the construction of individual confidence intervals for population means (these are of limited use because the confidence coefficient does not apply to a *series* of estimates). The power of ANOVA lies in providing us with the ability to make *joint* conclusions about population parameters.

As mentioned earlier, several procedures have been developed for further analysis. The method we will discuss here is the *Tukey method* of pairwise comparisons of

the population means. The method is also called the HSD (honestly significant differences) test. This method allows us to compare every possible pair of means using a *single level of significance*, say $\alpha = 0.05$ (or a single confidence coefficient, say, $1 - \alpha = 0.95$). The single level of significance applies to the *entire set* of pairwise comparisons.

The Tukey Pairwise Comparisons Test

We will use the *studentized range distribution*.

The **studentized range distribution**, q , is a probability distribution with degrees of freedom r and $n - r$.

Note that the degrees of freedom of q are similar, but not identical, to the degrees of freedom of the F distribution in ANOVA. The F distribution has $r - 1$ and $n - r$ degrees of freedom. The q distribution has degrees of freedom r and $n - r$. Critical points for q with different numbers of degrees of freedom for $\alpha = 0.05$ and for $\alpha = 0.01$ are given in Appendix C, Table 6. Check, for example, that for $\alpha = 0.05$, $r = 3$, and $n - r = 20$, we have the critical point $q_\alpha = 3.58$. The table gives right-hand critical points, which is what we need since our test will be a right-hand-tailed test. We now define the Tukey criterion, T .

The Tukey criterion

$$T = q_\alpha \frac{\sqrt{\text{MSE}}}{\sqrt{n_i}} \quad (8-21)$$

Equation 8-21 gives us a critical point, at a given level α , with which we will compare the computed values of test statistics defined later. Now let us define the hypothesis tests. As mentioned, the usefulness of the Tukey test is that it allows us to *jointly* perform all possible pairwise comparisons of the population means using a single, “family” level of significance. What are all the possible pairwise comparisons associated with an ANOVA?

Suppose that we had $r = 3$. We compared the means of three populations using ANOVA and concluded that the means were not all equal. Now we would like to be able to compare every *pair* of means to determine where the differences among population means exist. How many pairwise comparisons are there? With three populations, there are

$$\binom{3}{2} = \frac{3!}{2! 1!} = 3 \text{ comparisons}$$

These comparisons are:

- 1 with 2
- 2 with 3, and
- 1 with 3

As a general rule, the number of possible pairwise comparisons of r means is:

$$\binom{r}{2} = \frac{r!}{2!(r-2)!} \quad (8-22)$$

You do not really need Equation 8-22 for cases where it is relatively easy to list all the possible pairs. In the case of Example (b), there are, by Equation 8-22, $5!/2!3! = (5)(4)(3)(2)/(2)(3)(2) = 10$ possible pairwise comparisons. Let us list all the comparisons:

- Guadeloupe (1)–Martinique (2)
- Guadeloupe (1)–Haiti (3)
- Guadeloupe (1)–Paradise Island (4)
- Guadeloupe (1)–St. Lucia (5)
- Martinique (2)–Haiti (3)
- Martinique (2)–Paradise Island (4)
- Martinique (2)–St. Lucia (5)
- Haiti (3)–Paradise Island (4)
- Haiti (3)–St. Lucia (5)
- Paradise Island (4)–St. Lucia (5)

These pairings are apparent if you look at Table 8-4 and see that we need to compare the first island, Guadeloupe, with all four islands below it. Then we need to compare the second island, Martinique, with all three islands below it (we already have the comparison of Martinique with Guadeloupe). We do the same with Haiti, and finally with Paradise Island, which has only St. Lucia listed below it; therefore, this is the last comparison. (In the preceding list, we wrote the number of each population in parentheses after the population name.)

The parameter μ_1 denotes the population mean of all vacationer responses for Guadeloupe. The parameters μ_2 to μ_5 have similar meanings. To compare the population mean vacationer responses for every pair of island resorts, we use the following *set of hypothesis tests*.

- I. $H_0: \mu_1 = \mu_2$
 $H_1: \mu_1 \neq \mu_2$
- II. $H_0: \mu_1 = \mu_3$
 $H_1: \mu_1 \neq \mu_3$
- III. $H_0: \mu_1 = \mu_4$
 $H_1: \mu_1 \neq \mu_4$
- IV. $H_0: \mu_1 = \mu_5$
 $H_1: \mu_1 \neq \mu_5$

- V. $H_0: \mu_2 = \mu_3$
 $H_1: \mu_2 \neq \mu_3$
- VI. $H_0: \mu_2 = \mu_4$
 $H_1: \mu_2 \neq \mu_4$
- VII. $H_0: \mu_2 = \mu_5$
 $H_1: \mu_2 \neq \mu_5$
- VIII. $H_0: \mu_3 = \mu_4$
 $H_1: \mu_3 \neq \mu_4$
- IX. $H_0: \mu_3 = \mu_5$
 $H_1: \mu_3 \neq \mu_5$
- X. $H_0: \mu_4 = \mu_5$
 $H_1: \mu_4 \neq \mu_5$

The Tukey method allows us to simultaneously carry out all 10 hypothesis tests at a single given level of significance, say, $\alpha = 0.05$. Thus, if we use the Tukey procedure for reaching conclusions as to which population means are equal and which are not, we know that the probability of reaching at least one erroneous conclusion, stating that two means are not equal when indeed they are equal, is at most 0.05.

The test statistics:

The test statistic for each test is the *absolute difference of the appropriate sample means*.

Thus, the test statistic for the first test (I) is:

$$|\bar{x}_1 - \bar{x}_2| = |89 - 83| = 6$$

Conducting the Tests

We conduct the tests as follows. We compute each of the test statistics and compare them with the value of T that corresponds to the desired level of significance, α . We *reject a particular null hypothesis if the absolute difference between the corresponding pair of sample means exceeds the value of T* .

Using $\alpha = 0.05$, we now conduct the Tukey test for Example (b). All absolute differences of sample means corresponding to the pairwise tests I through X are computed and compared with the value of T . For $\alpha = 0.05$, $r = 5$, and $n - r = 195$ (we use ∞ , the last row in the table), we get, from Appendix C, Table 6: $q = 3.86$. We also know that $MSE = 504.4$ and $n_i = 40$ for all i . (Later we will see

what to do when not all r samples are of equal size.) Therefore, from Equation 8-21:

$$T = q_{\alpha} \sqrt{\frac{\text{MSE}}{n_i}} = 3.86 \sqrt{\frac{504.4}{40}} = 13.7$$

We now compute all 10 pairwise absolute differences of sample means and compare them with $T = 13.7$ to determine which differences are statistically significant at $\alpha = 0.05$ (these are marked with an asterisk).

$$|\bar{x}_1 - \bar{x}_2| = |89 - 75| = 14 > 13.7^*$$

$$|\bar{x}_1 - \bar{x}_3| = |89 - 73| = 16 > 13.7^*$$

$$|\bar{x}_1 - \bar{x}_4| = |89 - 91| = 2 < 13.7$$

$$|\bar{x}_1 - \bar{x}_5| = |89 - 85| = 4 < 13.7$$

$$|\bar{x}_2 - \bar{x}_3| = |75 - 73| = 2 < 13.7$$

$$|\bar{x}_2 - \bar{x}_4| = |75 - 91| = 16 > 13.7^*$$

$$|\bar{x}_2 - \bar{x}_5| = |75 - 85| = 10 < 13.7$$

$$|\bar{x}_3 - \bar{x}_4| = |73 - 91| = 18 > 13.7^*$$

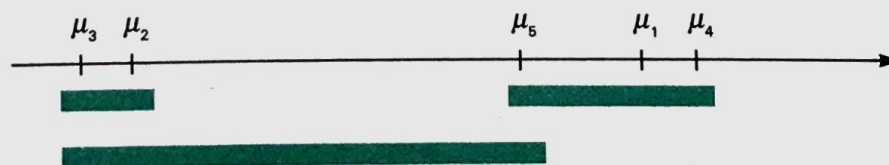
$$|\bar{x}_3 - \bar{x}_5| = |73 - 85| = 12 < 13.7$$

$$|\bar{x}_4 - \bar{x}_5| = |91 - 85| = 6 < 13.7$$

From these comparisons we determine that our data provide statistical evidence to conclude that μ_1 is different from μ_2 ; μ_1 is different from μ_3 ; μ_2 is different from μ_4 ; and μ_3 is different from μ_4 . *There are no other statistically significant differences at $\alpha = 0.05$.*

For the purpose of interpretation, it will help to draw a diagram of the significant differences that we found. This has been done in Figure 8-13 below. Looking at the figure, you may be puzzled by the fact that we believe, for example, that μ_1 is

FIGURE 8-13 Differences Among the Population Means in Example (b) Suggested by the Tukey Procedure



■ —Denotes that there is no statistical evidence that the covered population means are different

(The population means are placed approximately at the locations on the line where the sample means would be.)

different from μ_2 , yet we believe that μ_1 is no different from μ_5 and μ_5 is no different from μ_2 . You may say: If A is equal to B, and B is equal to C, then mathematically we must have A equal to C (the transitivity of equality). But remember that we are doing statistics, not discussing mathematical equality. In statistics, accepting the null hypothesis that two parameters are equal does not mean that they are necessarily equal. The acceptance just means that we have no statistical evidence to conclude that they are different. Thus, in our present example, we conclude that there is statistical evidence to support the claim that, on the average, vacationers give higher ratings to Guadeloupe (1) than they give to Martinique (2) or Haiti (3); as well as the claim that Paradise Island (4) is, on the average, rated higher than Martinique or Haiti. There is no statistical evidence for any other claim of differences in average ratings among the five island resorts. Note also that we do not have to hypothesize any of the assertions of tests I through X *before* doing the analysis. The Tukey method allows us to make all the above conclusions at a single level of significance, $\alpha = 0.05$.

The Case of Unequal Sample Sizes, and Alternative Procedures

What can we do if the sample sizes are not equal in all groups? We use the *smallest sample size* of all the n_i in computing the criterion T of Equation 8-21. The Tukey procedure is the best follow-up to ANOVA when the sample sizes are all equal. The case of equal sample sizes is called the *balanced design*. For very unbalanced designs (i.e., when sample sizes are very different), there are other methods of further analysis to be used following ANOVA. Two of the better-known methods are the Bonferroni method and the Scheffé method. We will not discuss these methods here.

Suppose that the same level of significance, α , is used for both the ANOVA and the ensuing Tukey analysis. If the ANOVA results in the rejection of the null hypothesis of equal population means, then the Tukey method (with a balanced design) guarantees that at least one significant difference will be found.