1 Chapter 3: Treatment Comparisons

Usually we have more specific research questions than just rejecting the ANOVA null hypothesis that all groups have the same mean ($H_0 : \mu_1 = \mu_2 = ... = \mu_t$).

Common questions:

1) While planning an experiment, you wish to test certain hypotheses that are a subset of the global ANOVA H_0. (*A priori* tests)

2) After a significant ANOVA, which groups differ? (*Post hoc* tests)

For example, the text discusses several questions that arise after finding that the ANOVA global $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected with the meat storage data.

1.1 Terminology

When describing treatment comparisons, it is useful to introduce some terminology. The term **contrast** is used to describe a comparison of means. Specifically, a contrast is a linear combination of the population means,

$$C = \sum_{i=1}^{t} k_i \mu_i \text{ that also satisfies } \sum_{i=1}^{t} k_i = 0.$$

We require that the coefficients k_i sum to zero so that the comparison is meaningful (we would not be interested in $\mu_1 - 3\mu_2$ for example). For example, if we had 6 treatment groups and we wished to test whether the the first two group means differed, we can express the null hypothesis as $H_0 : 1\mu_1 - 1\mu_2 = \mu_1 - \mu_2 = 0$. Here k_1=1 and k_2= -1, so $k_1 + k_2 = 0$ as required. This is an example of a **pairwise contrast**, which is defined as a contrast involving only two groups. An example of a **non-pairwise contrast** would be if we wished to test if the mean of the first two groups differed from the mean of the other four groups. We can express this null hypothesis as $H_0 : (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4 + \mu_5 + \mu_6)/4 = 0$. Here k_1=1/2, k_2= 1/2, k_3= -1/4, k_4= -1/4, k_5= -1/4 and k_6= -1/4, so again $k_1 + k_2 + k_3 + k_4 + k_5 + k_6 = 0$, as required by the definition of a contrast. One other property of a set of contrasts, called orthogonality, is useful when considering *a priori* tests. Two contrasts

$$C_1 = \sum_{i=1}^{t} k_{1i} \mu_i \text{ and } C_2 = \sum_{i=1}^{t} k_{2i} \mu_i \text{ are } \textbf{orthogonal} \text{ if } \sum_{i=1}^{t} \frac{k_{1i} k_{2i}}{n_i} = 0.$$

If the group sample sizes are equal then this is equivalent to $\sum k_{1i} k_{2i} = 0$. In the examples above, if we identify the first contrast as C_1 and the second contrast as C_2, then k_{11}=1, k_{12}= -1, $k_{13} = k_{14} = k_{15} = k_{16} = 0$, are the coefficients for C_1 and k_{21}=1/2, k_{22}= 1/2, k_{23}= -1/4, k_{24}= -1/4, k_{25}= -1/4 and k_{26}= -1/4 are the coefficients for C_2. Then if the sample sizes are equal,

$$\sum_{i=1}^{t} k_{1i} k_{2i} = (1)(1/2) + (-1)(1/2) + (0)(-1/4) + (0)(-1/4) + (0)(-1/4) + (0)(-1/4) = 0,$$

so C_1 and C_2 are orthogonal. Orthogonal contrasts are statistically independent, so that the outcome of testing one contrast is independent of the outcome of testing the other contrast. A set of more than two contrasts is **mutually orthogonal** if each pair of contrasts in the set is orthogonal to each other. The concept of a contrast or a set of contrasts at first seems somewhat esoteric, but in fact it is essential to

understand these concepts to fully understand ANOVA, particularly in complicated situations. The contrasts that you use should depend on the questions of scientific interest from your research.

2 Inference for contrasts

We can estimate the contrast

$$C = \sum_{i=1}^{t} k_i \mu_i$$

with the observed sample contrast:

$$c = \sum_{i=1}^{t} k_i \overline{y}_{i.}$$

The text presents formulas for the variance of a sample contrast that can be used to obtain a t test for the null hypothesis $H_0 : C = 0$.

2.1 Orthogonal contrasts can form a partition of SS Treatment

As noted in the text, we can form as many orthogonal contrasts as we have degrees of freedom for the treatment factor. These orthogonal contrasts partition the SS Treatment into SSC_i terms that allow us to separate the total sum of squares for treatments into parts attributable to different contrasts. Especially when you have equal sample sizes, this can be a powerful tool for understanding treatment effects. One common example of this is when we use orthogonal polynomial contrasts to partition quantitative treatment effects into parts attributable to linear trend, quadratic trend, and higher-order trends. For equally-spaced dosage levels with equal-sample-size groups, the coefficients for orthogonal polynomial contrasts are given in Table XI on page 623 of our text.

The SAS code for this lecture illustrates the use of orthogonal contrasts for the meat storage data, and also the use of the ORPOL function to obtain orthogonal polynomial coefficients for quantitative treatment levels. The code also shows the use of Proc Reg for obtaining predictions from response curves and standard errors of the predictions.

2.2 The need to control error rates with multiple comparisons

As shown in the text, when a set of n hypotheses are tested where each has a Type I error rate of α_C (the **comparisonwise error rate**), then an upper limit on the probability that at least one of the tests commits a Type I error (called the **experimentwise error rate**) is:

$$\alpha_E = 1 - (1 - \alpha_C)^n$$

Some values for these error rates are compared in Table 3.8 of the text, and they show that α_E increases rapidly as more tests are conducted.

2.3 A simple guide for multiple comparisons

There are a vast number of methods used for multiple comparison tests, and we will only consider a small number of them. We will only consider in detail three types of multiple comparison tests: **t tests**, **Tukey**'s method, and **Scheffe**'s method. **Fisher's LSD** method will also be discussed since it is widely used. The choice between these methods is governed by the type of contrasts being tested. In the somewhat artificial case in which a set of orthogonal contrasts has been specified *a priori*, then since the tests are independent we can simply apply separate t tests for each contrast, without adjusting the α level per contrast (this recommendation is at odds with the author of our text). If a set of contrasts has been specified *a priori* but are not orthogonal, t tests are again used but with a **Bonferroni correction**. In this case if n tests are involved, and the experimentwise error rate is to be held at α, then the comparisonwise error rate (for individual tests) is set at $\alpha' = \alpha/n$. Thus if 5 tests will be performed and the overall significance level for the set of tests is desired to be $\alpha = .05$, then $\alpha' = .05/5 = .01$ will be used for each individual test. If the contrasts to be tested are decided after collecting the data (*post hoc*) then we use generally more conservative methods to guard against data-snooping. For pairwise contrasts we can use Tukey's method and for non-pairwise contrasts we use Scheffe's method. This overall strategy is summarized in the following table, where the rows identify whether the contrasts are *a priori* or *post hoc*, and the columns identify whether they are orthogonal or not. Notice that all *post hoc* contrasts are treated as if they are nonorthogonal. The text considers some other alternative multiple comparison methods. If one treatment is considered a control, so that the pairwise contrasts of interest involve differences with the control, then **Dunnett's method** can be used. A related idea is to compare all treatments to the best (largest or smallest, depending on context) treatment, this is called the **multiple comparisons with the best** procedure. Some methods use multiple criteria to declare differences between treatment means based on the ordering of the sample means, these are called multiple range tests and the text presents the **Student-Newman-Keuls multiple range test**.

	Orthogonal	Nonorthogonal
A priori	Separate t-tests	Separate t-tests with Bonferroni correction
Post hoc		Pairwise: Tukey; Non-pairwise: Scheffe

2.4 A final note

As previously stated, there are a vast number of multiple comparison methods in use. We have only discussed a very small number of methods that are widely recognized, implemented in most software, and all provide confidence intervals as well as hypothesis tests. There are, for example, methods for pairwise contrasts that control Type I error for the collection of tests as Tukey's method does, but have much greater power for detecting differences. A good discussion of many methods is found in Chapter 4 of Kirk (1995). The methods discussed above were first developed for studies where a small or moderate number of comparisons are made. In some current scientific studies, much larger numbers of tests are made, such as in genomic studies. For these situations the concept of experimentwise error is usually no longer useful, and other criteria such as the false discovery rate (the expected proportion of false positives among the positive findings) are used instead.

3 Chapter 4: Model assumptions and transformations

3.1 Valid results depend on model assumptions

This point is discussed at the beginnnig of Chapter 4. Some general results on effects of departures from assumptions are mentioned, particularly problems that occur when variances are unequal and sample sizes are also unequal.

3.2 Model diagnostics: residual analysis

Recall that the residuals e_{ij} in the ANOVA model $y_{ij} = \mu_i + e_{ij}$ should 1) have a mean of 0, 2) be independent, 3) have a normal distribution, and 4) have a common variance σ^2. We can estimate the e_{ij} terms from our sample via one of several types of residuals:

Name of residual	Formula	SAS Proc GLM keyword	R function
residual	$\widehat{e}_{ij} = y_{ij} - \widehat{\mu}_{i.} = y_{ij} - \overline{y}_{i.}$	r = or residual =	residuals()
standardized residual	$w_{ij} = \widehat{e}_{ij}/\sqrt{MSE}$		
studentized residual	$\widetilde{e}_{ij} = \widehat{e}_{ij}/\sqrt{MSE(1 - 1/r_i)}$	student =	rstandard()
jackknife residual	$\widetilde{e}_{(-ij)} = \widehat{e}_{ij}/\sqrt{MSE_{(-ij)}(1 - 1/r_i)}$	rstudent =	rstudent()

where $MSE_{(-ij)}$ is MSE computed without the jth observation in the ith group. The last three types (standardized, studentized, and jackknife) are fairly similar if the model assumptions are satisfied. Since the last type, the jackknife residual, is best at detecting various problems, we will often use it. It can be obtained in SAS Proc GLM on the Output statement with the RSTUDENT option. A sound strategy for diagnosing the adequacy of a linear model is to obtain a variety of plots of residuals. These include histograms, normal plots, and scatter plots of the residuals versus the predicted values. A useful way to evaluate the collection of residual plots is to look at the plots for problems, and focus on characterizing the potential effect of these problems on the model. In addition to the various types of residuals shown above, the text discusses the use of the square root of the absolute residuals in a plot with the predicted values, and also some tests for homogeneity of variance, such as the Brown-Forsythe modification of the Levene test.

3.3 Transformations to help satisfy model assumptions

The text discusses different approaches for obtaining transformations of data to help meet model assumptions, focusing primarily on two methods, i) variance stabilizing transformations for known distributions, and ii) power transformations suggested by either a regression of $\log(\overline{y})$ on $\log(s)$ or use of the Box-Cox method.

The idea behind the first power transformation method is that if $\sigma_i = \alpha\mu_i^{\beta}$, then we have the relationship $\log(\sigma_i) = \log(\alpha) + \beta\log(\mu_i)$. The authors then show that if we select a power transformation $x = y^p$, then if p is chosen as $\widehat{p} = 1 - \widehat{\beta}$, the new variable x should have a variance that is approximately constant.

4 Reference

Kirk, R. E. (1995) Experimental Design: Procedures for the Behavioral Sciences. Pacific Grove: Brooks/Cole.