Usual Analysis of variance procedure

1) H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$

2) H_1: some μ_i is different

3) a) Assume that the observations are normally distributed about each mean, or that the residuals (i.e. deviations) are normally distributed.

b) Assume that the observations are independent

c) Assume that the variances are homogeneous

4) Set the level of type I error. Usually $\alpha = 0.05$

5) Determine the critical value. The test is in ANOVA is a one tailed F test.

6) Obtain data and evaluate.

7) Conclusions

Descriptions of post-hoc tests

Post-hoc or Post-ANOVA tests! Once you have found out some treatment(s) are "different", how do you determine which one(s) are different?

If we had done a t-test on the individual pairs of treatments, the test would have been done as $t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{MSE\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$. If the difference between $\bar{Y}_1 - \bar{Y}_2$ was large enough, the t value would have been greater than the $t_{critical}$ and we would conclude that there was a significant difference between the means. Since we know the value of $t_{critical}$ we could figure out how large a difference is needed for significance for any particular values of MSE, n_1 and n_2. We do this by replacing t with $t_{critical}$ and solving for $\bar{Y}_1 - \bar{Y}_2$.

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{MSE\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \text{ so}$$

$$t_{critical}\sqrt{MSE\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \bar{Y}_1 - \bar{Y}_2 \quad \text{or} \quad \bar{Y}_1 - \bar{Y}_2 = t_{critical} S_{\bar{Y}_1 - \bar{Y}_2}$$

This value is the exact width of an interval $\bar{Y}_1 - \bar{Y}_2$ which would give a t-test equal to $t_{critical}$. Any larger values would be "significant" and any smaller values would not. This is called the "Least Significant Difference". $\boxed{LSD = t_{critical} S_{\bar{Y}_1 - \bar{Y}_2}}$

This least significant difference calculation can be used to either do pairwise tests on observed differences or to place a confidence interval on observed differences.

The LSD can be done in SAS in one of two ways. The MEANS statement produces a range test (LINES option) or confidence intervals (CLDIFF option), while the LSMEANS statement gives pairwise comparisons.

The LSD has an α probability of error on each and every test. The whole idea of ANOVA is to give a probability of error that is α for the whole experiment, so, much work in statistics has been dedicated to this problem. Some of the most common and popular alternatives are discussed below. Most of these are also discussed in your textbook.

The LSD is the LEAST conservative of those discussed, meaning it is the one most likely to detect a difference and it is also the one most likely to make a Type I error when it finds a difference. However, since it is unlikely to miss a difference that is real, it is also the most powerful. The probability distribution used to produce the LSD is the t distribution.

Bonferroni's adjustment. Bonferroni pointed out that in doing k tests, each at a probability of Type I error equal to α, the overall experimentwise probability of Type I error will be NO MORE than $k*\alpha$, where k is the number of tests. Therefore, if we do 7 tests, each at α=0.05, the overall rate of error will be NO MORE than =.35, or 35%. So, if we want to do 7 tests and keep an error rate of 5% overall, we can do each individual test at a rate of $\alpha/k = 0.055/7 = 0.007143$. For the 7 tests we have an overall rate of $7*0.007143 = 0.05$. The probability distribution used to produce the LSD is the t distribution.

Duncan's multiple range test. This test is intended to give groupings of means that are not significantly different among themselves. The error rate is for each group, and has sometimes been called a familywise error rate. This is done in a manner similar to Bonferroni, except the calculation used to calculate the error rate is $[1-(1-\alpha)^{r-1}]$ instead of the sum of α. For comparing two means that are r steps apart, where for adjacent means r=2. Two means separated by 3 other means would have r = 5, and the error rate would be $[1-(1-\alpha)^{r-1}] = [1-(1-0.05)^4] = 0.1855$. The value of a needed to keep an error rate of α is the reverse of this calculation, $[1-(1-0.05)^{1/4}] = 0.0127$.

Tukey's adjustment The Tukey adjustment allows for **all possible pairwise tests**, which is often what an investigator wants to do. Tukey developed his own tables (see Appendix table A.7 in your book, "percentage points of the studentized range). For "t" treatments and a given error degrees of freedom the table will provide 5% and 1% error rates that give an experimentwise rate of Type I error.

Scheffé's adjustment This test is the most conservative. It allows the investigator to do not only all pairwise tests, but **all possible tests**, and still maintain an experimentwise error rate of α. "All possible" tests includes not only all pairwise tests, but comparisons of all possible combinations of treatments with other combinations of treatments (see CONTRASTS below). The calculation is based on a square root of the F distribution, and can be used for range type tests or confidence intervals. The test is more general than the others mentioned, for the special case of pairwise comparisons, the statistic is $\sqrt{(t-1)*F_{t-1,\, n(t-1)}}$ for a balanced design with t treatments and n observations per treatment.

Place the post-hoc tests above in order from the one most likely to detect a difference (and the one most likely to be wrong) to the one least likely to detect a difference (and the one least likely to be wrong). **LSD is first, followed by Duncan's test, Tukey's and finally Scheffé's**. Dunnett's is a special test that is similar to Tukey's, but for a specific purpose, so it does not fit well in the ranking. The Bonferroni approach produces an upper bound on the error rate, so it is conservative for a given number of tests. It is a useful approach if you want to do a few tests, fewer than allowed by one of the others (e.g. you may want to do just a few and not all possible pairwise). In this case, the Bonferroni may be better.

Contrasts

A calculation similar to the LSD, but extended to more than just 2 means, is called a contrast. Suppose we wish to test the mean of the first two means against the mean of the last 3 means.

1) H_0: $\frac{\mu_1+\mu_2}{2}=\frac{\mu_3+\mu_4+\mu_5}{3}$ or $\frac{\mu_1+\mu_2}{2}-\frac{\mu_3+\mu_4+\mu_5}{3}=0$ or

$$\left(\frac{1}{2}\mu_1+\frac{1}{2}\mu_2\right)-\left(\frac{1}{3}\mu_3+\frac{1}{3}\mu_4+\frac{1}{3}\mu_5\right)=0 \text{ or}$$

$$\frac{1}{2}\mu_1+\frac{1}{2}\mu_2+\left(-\frac{1}{3}\right)\mu_3+\left(-\frac{1}{3}\right)\mu_4+\left(-\frac{1}{3}\right)\mu_5=0 \text{ or}$$

$$3\mu_1+3\mu_2+(-2)\mu_3+(-2)\mu_4+(-2)\mu_5=0$$

This expression is what is a "linear model", and the last expression of this linear model is the easiest form for us to work with. We can evaluate the linear model, and if we can find the variance we can test the linear model. Generically, the variance of a linear model is "the sum of the variances", however there are a few other details. As with the transformations discussed earlier in the semester, when we multiply a value by "a" the mean changes by "a", but the variance changes by "a^2". Also, if there are covariances between the observations these must also be included in the variance. For our purposes, since we have assumed independence, there are no covariances.

The linear expression to evaluate is then: $a_1T_1+a_2T_2+a_3T_3+a_4T_4+...+a_kT_k$ where the "a" are the coefficients and the "T" are the treatment means (sums can also be used).

The variance is then: $a^2{}_1\text{Var}(T_1)+a^2{}_2\text{Var}(T_2)+a^2{}_3\text{Var}(T_3)+a^2{}_4\text{Var}(T_4)+...+a^2{}_k\text{Var}(T_k)$

In an ANOVA, the best estimate of the variance is the MSE, and the variance of a treatment mean is MSE/n, where n is the number of observations in that treatment. We can therefore factor out MSE, and in the balanced case (1/n) can also be factored out. The result is $(MSE/n)(a^2{}_1+a^2{}_2+a^2{}_3+a^2{}_4+...+a^2{}_k)$

If we were to use a t-test to test the linear combination against zero, the t-test would be:

$$\frac{a_1T_1+a_2T_2+a_3T_3+a_4T_4+\ldots+a_kT_k}{\sqrt{\frac{MSE}{n}\left(a_1^2+a_2^2+a_3^2+a_4^2+\ldots+a_k^2\right)}}=\frac{\sum_{i=1}^{k}a_iT_i}{\sqrt{\frac{MSE}{n}\sum_{i=1}^{k}a_i^2}}$$

This is the test done with treatment means. If treatment totals are used the equation is modified slightly to $\sum_{i=1}^{k}a_iT_i\Big/\sqrt{nMSE\sum_{i=1}^{k}a_i^2}$ and will give the same result.

One final modification. If we calculate our "contrasts" as above without the "MSE" in the denominator, then we calculate $Q=\sum_{i=1}^{k}a_iT_i\Big/\sqrt{n\sum_{i=1}^{k}a_i^2}$, without the MSE, then all that would remain to complete the t-test is to divide by $\sqrt{MSE}$.

The value called "Q", when divided by $\sqrt{MSE}$ gives a t statistic. If we calculate Q^2 and divide by MSE we get an F statistic. SAS uses F tests. All we need provide SAS is the values of "a", the coefficients, in the correct order, and it will calculate and test the "Contrast" with an F statistic.