Adjustments for Multiple Testing and Estimation

If we conduct many tests, we are likely to see some small p-values even if the null hypothesis is true for every test. Suppose the null hypothesis in true for each of m independent tests.

$$P(\text{at least one } p\text{-value less than } 0.05) = 1 - P(\text{no } p\text{-value less than } 0.05) = 1 - 0.95^m$$

m	1	2	3	$\cdots$	10	$\cdots$	100
$1 - 0.95^m$	0.0500	0.0975	0.1426	$\cdots$	0.4013	$\cdots$	0.9941

Similarly if we construct many 95% confidence intervals, the chance that they all contain the true values of the parameters that they estimate will be lower than 95%. For m independent confidence intervals we have

$$P(\text{all confidence intervals cover their parameters}) = 0.95^m.$$

m	1	2	3	$\cdots$	10	$\cdots$	100
0.95^m	0.9500	0.9025	0.8574	$\cdots$	0.5987	$\cdots$	0.0059

The *individual confidence level* is the frequency with which a single interval captures its parameter. The *familywise confidence level* is the frequency with which all intervals in a family (i.e. group of several intervals) simultaneously capture their parameters. The *Bonferroni correction* is one method of ensuring that the familywise confidence level will be at least $100(1-\alpha)\%$. Bonferroni-corrected confidence intervals are obtained by using $t_{df}^{(1-\alpha/(2m))}$ in place of $t_{df}^{(1-\alpha/2)}$ when constructing m confidence intervals. The probability that all the Bonferroni-corrected intervals will contain their target parameters simultaneously is greater than or equal to $1-\alpha$.

The Bonferroni correction can also be used to keep the probability of rejecting one or more true null hypotheses below α. Bonferroni-corrected p-values are obtained by multiplying each p-value by the number of tests conducted. (Any product greater than 1 is set to 1.) This is a conservative procedure that will favor accepting the null hypothesis. When Bonferroni-corrected p-values are compared to 0.05 to determine significance, the chance of one or more rejections of a true null hypothesis is less than or equal to 0.05.

1. An experiment was conducted to compare the effectiveness of 5 chemical treatments (A, B, C, D, and E) at reducing damage caused by a natural pathogen. 30 pots, each containing two plants, were assigned to the 5 chemical treatments according to a completely randomized design. For each of the 6 pots randomly assigned to a particular chemical treatment, one of the plants (selected at random) was treated with the chemical while the other was treated with a placebo. All plants were infected with the pathogen. After 3 days, the number of lesions on the placebo-treated plant minus the number of lesions on the chemical-treated plant was determined for each pot. Use the results from the table below to produce a confidence intervals for each of the five chemical means that have familwise confidence level at least 95%. Based on your confidence intervals, which chemical treatments appear to be effective at reducing the number of lesions?

Chemical	**Mean of Differences**	**Variance of Differences**
A	1.6	4.3
B	2.1	4.8
C	5.3	3.3
D	7.7	3.6
E	8.6	4.0

There are many methods for addressing the multiple testing and estimation issue that are more sophisticated than the Bonferroni method. We will look briefly at Tukey's method for controlling the familywise confidence level when examining all possible differences between two means in the one-way ANOVA setting. In one-way ANOVA there are I group means. There are $\binom{I}{2}$ possible comparisons of two means. We could construct $\binom{I}{2}$ Bonferroni-corrected confidence intervals for the difference between two means. We could declare significant differences between any pair of means whose confidence interval does not include zero. Equivalently we could conduct $\binom{I}{2}$ tests and multiply the p-values by $\binom{I}{2}$ to determine which differences are significant. These methods would control the familywise confidence level. Tukey's method also controls the familywise confidence level in this situation and is less conservative than the Bonferroni method. Tukey's method produces intervals that contain the true difference between each pair of means simultaneously with probability $1 - \alpha$. Confidence intervals that do not contain zero correspond to Tukey-corrected p-values less than 0.05. The probability of rejecting one or more true null hypotheses is no larger than α when using the Tukey method to judge significance. When the within-group samples sizes $n_1, n_2, \ldots, n_I$ are not all equal, the Tukey-Kramer procedure will give an approximate familywise confidence level of $100(1 - \alpha)\%$.

When comparing pairs of means, a value known as the *minimum significant difference* is often computed. Any difference between means greater than or equal to the minimum significant difference is considered to be a statistically significant difference, while differences less than the minimum significant difference are considered to be nonsignificant. The minimum significant difference is also the number to the right of the $\pm$ in a confidence interval for the difference between a pair of means. A minimum significant difference computed using the Tukey method is also known as an *honestly signficant difference* (HSD).

2. The HSD corresponding to 95% confidence is 3.40 for the data in problem 1. Use this HSD to determine which pairs of means differ from one another in problem 1.

3. A *line plot* can be used to exhibit significant differences in the one-way ANOVA setting when sample sizes are equal. The means for each treatment are plotted labeled on a number line. A line is drawn above the number line to connect any means that are not significantly different from one another. Make a line plot for the data from problem 1.