

Adjustments for Multiple Testing and Estimation

If we conduct many tests at the 0.05 level, we are likely to see some significant results even if the null hypothesis is true for every test.

$$P(\text{at least one Type I error in } m \text{ independent tests}) = 1 - P(\text{no Type I errors}) = 1 - 0.95^m$$

m	1	2	3	...	10	...	100
$1 - 0.95^m$	0.0500	0.0975	0.1426	...	0.4013	...	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	...	10	...	100
0.95^m	0.9500	0.9025	0.8574	...	0.5987	...	0.0059

Bonferroni Corrections for Multiple Tests and Confidence Intervals

The *Bonferroni Correction* is one method of ensuring that the probability of one or more Type I errors (a.k.a. *the overall or experimentwise type I error rate*) is controlled. The correction can also be used to ensure that all confidence intervals will simultaneously contain their target parameters with 95% confidence. Bonferroni-corrected P-values are obtained by multiplying each P-value by the number of tests conducted. 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 type I errors (the overall type I error rate) is less than or equal to 0.05.

Bonferroni-corrected confidence intervals are obtained by using $t_{0.025/m}$ in place of $t_{0.025}$ when constructing m 95% confidence intervals. The probability that all the Bonferroni-corrected intervals will contain their target parameters simultaneously is greater than or equal to 0.95.

Tukey's Method for All Pairwise Comparisons

In one-way ANOVA there are t group means. We want to know which means, if any, are different from each other. There are $\binom{t}{2}$ possible comparisons of two means. We could construct $\binom{t}{2}$ 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 $\binom{t}{2}$ tests and multiply the P-values by $\binom{t}{2}$ to determine which differences are significant. These methods would control the overall type I error rate. Tukey's method also controls the overall type I error rate in this situation and is less conservative than the Bonferroni method. Tukey confidence intervals for the difference between all pairs of means are obtained by using SAS code like that below.

```
proc glm;
  class group;
  model y=group;
  means group / tukey cldiff alpha=0.05;
```

Here y and $group$ are the names of variables that will change from problem to problem. The *cldiff* term asks for confidence intervals for the difference between each pair of means. These intervals will contain the true difference between each pair

of means simultaneously with probability $1 - \alpha$. In this example the value of α is specified to be 0.05 using the command *alpha=0.05*. Confidence intervals that do not contain zero indicate a significant difference between means. The overall probability of a type I error is no larger than α when using the Tukey method to judge significance. This is true regardless of the true values of the treatment means.

Hsu's Method for All Possible Comparisons with the Largest Observed Mean

Suppose the goal of our analysis is to determine the treatment that has the largest mean μ_i . We know that the treatment with the largest mean μ_i will not always produce the largest sample mean \bar{y}_i . After all \bar{y}_i is only an estimate of μ_i that is subject to sampling variability. In many cases we will not be able to say that one particular treatment has a mean that is statistically significantly larger than all the others. Often we will be able to partition the treatments into two groups : (i) a set of treatments whose observed means are not significantly smaller than the largest observed mean and (ii) a set of treatments whose observed means are significantly smaller than the largest observed mean. If we conduct our analysis at the 0.05 level, the set of treatments in (i) will contain the treatment with the largest mean with probability 0.95.

Hsu's method for determining the sets (i) and (ii) works by obtaining simultaneous confidence intervals for the difference between the treatment mean corresponding to the largest observed mean and each of the other treatment means. We will illustrate the method by using the example referred to in Section 3.6 of Kuehl's text. Hsu's confidence intervals can be obtained using SAS Version 8.1 by following the instructions below.

Run the SAS program *filter.sas* (available on the Stat402A web site). This program reads in data on filter type and flow rate. Click on the *Solutions* menu at the top of the SAS screen. Select *Analysis* and then *Interactive Data Analysis*. Under *Library*: highlight *Work*. Under *Data Set*: highlight *FILTERDATA*. Click *Open* at the bottom of the dialogue box. You should see the data set pop up in a spreadsheet-like window. Now click on the *Analyze* menu at the top of the SAS screen (if running SAS for Windows) or at the top of the spreadsheet window (if running SAS on Project Vincent). Select *Box Plot/Mosaic Plot(Y)*. You should see the variable names *filter* and *flowrate* in a dialogue box. Click on *filter* and then the *X* button in the right portion of the dialogue box. The variable name *filter* should now appear below the *X* button. You have just designated *filter* as the explanatory variable. Now click on *flowrate* and the *Y* button in the right portion of the dialogue box. The variable name *flowrate* should now appear below the *Y* button. You have just designated *flowrate* as the response variable. Now click the *Output* button at the bottom of the dialogue box. Click in the small box immediately to the left of *Multiple Comparison of Means*. Click the *Multiple Comparison Options* button. Click the circle or diamond to the left of *Hsu's Test for Best*. Click *OK* and *OK* a second time in the next dialogue box. Finally click *OK* a third time. You should see side-by-side box plots for comparing the flow rates of the 7 filter types. By expanding the viewable region of the plot you should see output similar to that found in Table 3.9 on page 101 of Kuehl's text. One difference is that SAS doesn't report an interval for the largest of the observed sample means because that information is redundant. (The interval for the largest is negative one times the interval for the second largest.)

Confidence intervals with 0 as their upper endpoint correspond to treatment means that are significantly smaller than the largest of the observed means. These treatments go in set (ii) described previously. Intervals that properly contain 0 correspond to treatment means that are not significantly different from the largest mean. These treatments, along with the treatment that had the largest observed mean, make up set (i). In the filter example, we would say that either filter A or filter E has the largest mean. None of the other filters have the largest mean. These statements hold true with 95% confidence.

There is, of course, an analogous method for making all possible comparisons with the smallest observed mean. You should be able to conduct this analysis by making the obvious adjustments to the method described above for making all possible comparisons to the largest observed mean.

Comparison of All Treatments with a Control

Sometimes the goal of the analysis is to determine the set of treatments with means that are different from the control mean. This can be accomplished in SAS using the *dunnnett* option for the *means* statement in *proc glm*. Consider again the data on filters and flow rates. As in Section 3.7 of Kuehl's text, we assume that filter F is the control. The code contained in *filter.sas* will provide simultaneous confidence intervals for the differences between each treatment (A through E) and the control F. The probability that all such confidence intervals will contain the true differences between each treatment mean and the control mean is no smaller than the specified $1 - \alpha$ level, regardless of the values of the true means. One-sided tests can be obtained by replacing *dunnnett* with *dunnnett1* (H_A : treatment < control) or *dunnnettu* (H_A : treatment > control).