

Multiple Testing/Estimation

Idea: In the ANOVA setting, sometimes we have many CIs & tests at one time (i.e., to estimate μ_i with a CI or to compare μ_i & μ_j with a test). When we do a lot of tests or CIs, we need to make **adjustments**.

To see why, consider an example with CIs:

- Suppose we have m independent CIs, each CI has an **individual confidence level** of (say) 95%.¹
- For a large number m of CIs, there's a high chance that not all m CIs will contain the parameters that they estimate.

e.g. out of $m = 100$ 95% CIs, how many CIs would you expect to contain their parameters?

- To summarize:
 - \Rightarrow for m CIs, we want to make these CIs so that the **familywise confidence level** is high or, in other words,
 - \Rightarrow we want high confidence that all m CIs simultaneously (as a family) contain the parameters that they are respectively trying to estimate
 - e.g., want 95% (or higher) confidence that *simultaneously all m CIs* contain their parameters, not just 95% confidence for any individual interval

Note: We discussed problems with multiple CIs above, but similar problems exist in conducting a large number m of hypothesis tests. That is, if we test m hypotheses *and all m are true*, then we can expect some tests to have a small p-value just by chance (because of numerous tests). So, as with CIs, we have to make adjustments when there are many tests.

¹That is, before we collect data, for any single interval there is a 95% chance that the parameter (e.g., μ_i or $\mu_1 - \mu_2$) will fall in the CI.

Two methods for multiple tests and CIs:

1. Bonferroni

Bonferroni CIs (general method, doesn't depend on the CI):

To make m CIs with (at least) $100(1 - \alpha)\%$ familywise confidence, use $t_{df}^{1-\alpha/(2m)}$ instead of $t_{df}^{1-\alpha/2}$

Bonferroni Tests (general method, doesn't depend on tests):

To make m hypothesis tests,

- a) get p-values p_1, \dots, p_m from m single tests
- b) multiply these by m for "new" p-values: $p_1^{new} = mp_1, \dots, p_m^{new} = mp_m$
- c) treat $p_1^{new}, \dots, p_m^{new}$ as the p-values for interpreting each test as usual.

2. Tukey (for *all comparisons* $\mu_i - \mu_j$ between two groups)

Tukey CIs

To make CIs with familywise confidence of $100(1 - \alpha)\%$ for all possible mean differences $\mu_i - \mu_j$, use $\bar{Y}_i - \bar{Y}_j \pm HSD$

Notes:

1. The value HSD is called the honest significant difference (you do not compute this by hand, but use software and use it like a margin of error).
2. The familywise confidence level is $100(1 - \alpha)\%$, meaning you can be $100(1 - \alpha)\%$ confident that each $\mu_i - \mu_j$ falls into its CI $\bar{Y}_i - \bar{Y}_j \pm HSD$ for *all* comparisons.
3. Two population/treatment means μ_i & μ_j are declared different if 0 is not in the CI: $\bar{Y}_i - \bar{Y}_j \pm HSD$, or equivalently if $|\bar{Y}_i - \bar{Y}_j| \geq HSD$ holds.

Tukey Tests

Test $H_0 : \mu_i = \mu_j$ over all $m = I(I - 1)/2$ possible comparisons of two groups i, j

The Tukey method will give p-values that are adjusted for testing all these hypotheses $H_0 : \mu_i = \mu_j$ at the same time. SAS will do this.