Worksheet 08

1. Consider running four hypothesis tests that result in the following raw p-values: $0.001, 0.015, 0.2, 0.8$. What are adjusted p-values using the Bonferroni correction? How many tests are significant at the 0.05 level after the correction?[1]

[1] If the adjusted value is greater than 1, we usually set it to 1 as the defintion of a p-value is given as a probability at we can safely always use 1 as an upper bound.

2. The FWER only considers whether we have made at least one false rejection of one of the hypothesis tests. It does not matter, at least for FWER, if we make one mistake or many. Thinking about this for a moment, we might consider setting all of the p-values equal to the smallest p-value multiplied by the number of tests. Why? Let's consider just two tests for hypotheses H_1 and H_2 with p-values p_1 and p_2. Consider adjusting each of the p-values to be $2 \times \min(p_1, p_2)$. (a) If both H_1 and H_2 are true, what is the (upper bound) on the FWER if we use the adjusted p-values at a level α? (b) If neither H_1 and H_2 are true, what is the FWER? (c) What can we say about the FWER if H_1 is true but not H_2 (or vice-versa)?

3. Consider an adjusted version of the above procedure. We adjust the smallest p-value to be twice its original value, but set the larger p-value to be the smaller of the adjusted smaller value or the original second largest p-value (it sounds more complex than it is). (a) If both H_1 and H_2 are true, what is the (upper bound) on the FWER if we use the adjusted p-values at a level α? (b) If neither H_1 and H_2 are true, what is the FWER? (c) What can we say about the FWER if H_1 is true but not H_2 (or vice-versa)?

4. The **Holm-Bonferroni** correction is a generalization of the above for any number of tests. Consider a set of m sorted p-values from smallest to largest: $p_1, \ldots, p_m$. We adjust them according to the following iterative procedure (setting $p'_0 = 0$):

$$p'_j = \max\{p'_{j-1}, \frac{p_j}{m+1-j}\}$$

This will control the FWER at the level α for the same logic that we derived above in the two-test case. What are the Holm-Bonferroni corrected p-values from the first question? How many tests are significant at the 0.05 level after the correction?