**PSTAT 5A: Discussion Worksheet 05**

*Summer Session A 2023*, with *Ethan P. Marzban*

1. Ecologists near the remote village of *Statsville* in the nation of *Gauchonia* would like to determine whether the village's drinking water has the same amount of bacteria as drinking water found near the capital city. To that end, they take 15 samples of water near *Statsville* and 20 samples of water near the capital, and record the bacteria levels (in cells per liter) in each sample. The ecologists' results are summarized below:

|  | Sample Mean | Sample Std. Dev. |
|---|---|---|
| **Statsville** | 75 | 12.3 |
| **Capital City** | 63 | 18.7 |

   Let "Population 1" refer to the drinking water in *Statsville* and "Population 2" refer to the drinking water near the capital city. Additionally, assume all independence and normality conditions are satisfied; also use a 5% level of significance and a two-sided alternative wherever necessary.

   (a) Define the parameters of interest, $\mu_1$ and $\mu_2$.

   (b) State the null and alternative hypotheses.

   (c) Compute the observed value of the test statistic.

   (d) Assuming the null is correct, what is the approximate distribution of the sampling distribution? Be sure to include any/all relevant parameters.

   (e) What is the $p-$value of the observed test statistic? (As a reminder, you may need to use Python for this part.)

   (f) What is the critical value of the test?

   (g) Now, carry out the test. Be sure to phrase your conclusions in terms of the context of the problem.

   (h) Re-do the test, now using the alternative that the capital city has higher bacterial levels in its water than *Statsville*. Again use a 5% level of significance.

   (i) Re-do the test, now using the alternative that *Statsville* has higher bacterial levels in its water than the capital city. Again use a 5% level of significance.

2. The following ANOVA table has some entries missing. Fill in the missing entries, and provide justification as to how you found those missing values. You may need Python for certain entries.

|  | DF | Sum Sq. | Mean Sq. | $F$−value | $\mathbb{P}(> F)$ |
|---|---|---|---|---|---|
| **Btw. Groups** | 12 | 18 | 1.5 | <???> | 0.02976 |
| **Residuals** | 120 | <???> | <???> |  |  |