# MATH 189: Homework 2

## Due Jan 27th 2023

The *USDA Women's Health Survey* dataset (nutrient.txt) contains five types of women's nutrient intakes which were measured from a random sample of 737 women aged 25-50 years in United States. Analyze the dataset according to the following steps:

1. Calculate sample mean and sample standard deviation of each variable.

2. The recommend intake amount of each nutrient is given in the following table. For each nutrient, apply a univariate t-test to test if the population mean of that variable equals the recommended value. Set the significance level at $\alpha = 0.05$.

| Variable | Calcium | Iron | Protein | Vitamin A | Vitamin C |
|---|---|---|---|---|---|
| Recommended Intake Amount | 1000mg | 15mg | 60g | 800$\mu$g | 75mg |

3. Based on the results you obtained in step 2, how would you interpret your test results? Do you think the US Women meet the recommended nutrient intake amount? If not, what would you suggest to the public?

## Problem 2

The *Multiple Testing* dataset (multiple.txt) is a simulated dataset which contains 50 variables and 100 observations per variable. Suppose we know that the first 10 variables have mean equal to 2 and the rest of them have mean equal to 0. Analyze the dataset according to the following steps:

1. Perform multiple testing to the population mean vector to test if it equals to a vector whose elements are all zeros. Set the significance level at $\alpha = 0.1$ .

2. Based on the test results in step 1, calculate the following quantities: number of type I errors, FWER and FDP.

3. Redo the multiple testing in step 1 with Bonferroni correction (set $\alpha = 0.1$). Calculate the FWER of your new test results.

4. Redo the multiple testing in step 1 with BH procedure (set $\alpha = 0.1$). Calculate the FDP and FWER of your new test results. How does the results compared with the ones you obtained in step 1 and step 3?