**Lab Assignment 2: Statistical Analysis (due Feb 9)**

**Instructions**

* Load GACTT\_RESULTS\_ANONYMIZED\_HW2.RDS into R.
* Ensure your results are **clear, well-organized, and properly interpreted** with necessary visualizations.
* Where applicable, **provide descriptive statistics** before running formal tests.
* For all statistical analyses, **state your hypotheses** in math and **interpret the results in context**.

**1. Data Preparation**

* The outcome variable for this assignment is **a\_preference** (revealed preference for a tasting lot for Coffee type A, which is a light-roasted coffee from Kenya).
* Missing values in a\_preference are **coded as -9**. Recode them to NA.
* Consider whether dropping missing values is appropriate for your analysis.

**2. t-test: Coffee Preparation Method & Preference (3 pts)**

* Investigate whether **people whose favorite drink is "pour over"** differ in preference for the tasting lot compared to those who prefer other drinks.
* Conduct an **independent samples t-test** to compare **a\_preference** between these two groups.
* Provide appropriate **visualization**. Combine the visualization with your test result in your **interpretation**:
  + Are the differences statistically significant?
  + Do people who prefer pour-over coffee **enjoy the tasting lot more** than others?

**3. ANOVA: Coffee Strength & Preference (3 pts)**

* Work with the **preferred coffee strength** variable (preferred\_strength), currently coded from **1 (very light) to 5 (very strong)**.

**Step 1: Recode preferred\_strength**

* Create a new categorical variable with three levels:
  + **"Strong"** (values 4 & 5)
  + **"Medium"** (value 3)
  + **"Light"** (values 1 & 2)
* Retain the original numeric variable in the dataset.

**Step 2: Run One-Way ANOVA**

* Conduct a **one-way ANOVA** using preferred\_strength (categorical) to predict a\_preference.
* Provide appropriate visualization
* **Interpretation:**
  + What is the general perception of the tasting lot among different strength preference groups?
  + If significant, which groups differ from each other?

**4. OLS Regression: Comparing Models (4 pts)**

Conduct two **Ordinary Least Squares (OLS) regressions** and compare results to previous tests.

**Step 1: Regression with "Pour Over" Preference (2 pts)**

* Return to the **"pour over" vs other drinks** comparison.
* Regress a\_preference on the **"pour over" indicator variable** (same predictor as in **Question 2**).
* Compare the regression results to the **t-test results**. Are they consistent?

**Step 2: Regression with Coffee Strength (2 pts)**

* Use the **numeric version** of preferred\_strength as a predictor for a\_preference.
* Compare the regression results to the **ANOVA results** from **Question 3**.
* Which method (ANOVA or regression) is **more appropriate** for this task? Why?

**5. Bonus Task: Custom t-test Function (2 pts)**

**Step 1: Write a Function**

* Write an **R function** to perform a **one-sample t-test**.
* The function should:
  + Perform **both one-tailed and two-tailed tests**.
  + Allow the user to specify a **population mean** for comparison.
  + **Do not use the built-in t.test() function** in your implementation.

**Step 2: Test Your Function**

* Use your function on **multiple test cases** to verify its correctness.
* Consider **edge cases** (e.g., very small sample sizes, missing values, extreme values).
* Identify scenarios where the function might fail or return unexpected results.