Question 1 *(20 points)*

The officers at the Internal Revenue Service (IRS) in the United States and Canada Revenue Agency (CRA) are always looking for ways to improve the wording and format of its tax return forms. Three new forms have been developed recently. To determine which, if any, are superior to the current form, 120 individuals were asked to participate in an experiment. Each of the three new forms and the currently used form were filled out by 30 different people. The amount of time (in minutes) taken by each person to complete the task was recorded and are given in the CSV file *IRS 1*. Use a 5% significance level to answer the following questions.

a) Assess whether or not the normality condition is satisfied.

b) Assess whether or not the equal variances condition is met by running the test of equal variances.

c) Run a one-way ANOVA and comment on whether or not differences exist among the forms.

d) If differences exit, use Tukey's test to identify which forms are different.

Question 2 *(35 points)*

Suppose the experiment in Question 1 is redone in the following way. Thirty people are asked to fill out all four forms. The completion times (in minutes) are recorded and provided in the CSV file *IRS 2*. Use a 5% significance level to answer the following questions.

a) Assess whether or not the normality condition is satisfied. (If the condition appears to be not satisfied, then at this time, ignore this requirement and still assume it is satisfied.)

b) Assess whether or not the equal variances condition is met by running the test of equal variances.

c) Run an appropriate ANOVA test and comment on whether or not differences exist among the forms.

d) Comment on the suitability of the experimental design chosen in part c above. Was blocking effective?

e) Use the Tukey's test to identify which forms are different.

f) Compare the results of the Tukey test from Questions 1 and 2 and comment on why you think the results are different.

g) In the ANOVA table, the `Error` value under `Adj SS` column shows the variation **within** each treatment group. This measure is denoted by `SSE (Sum of Squares Error)`. Note the values from the R ANOVA tables of Questions 1 and 2 and comment on why you think the `SSE` value in Question 2 is lower.

Question 3 *(18 points)*

Suppose that the experiment in Questions 1 and 2 is redone in the following way. Thirty taxpayers fill out each of the four forms. However, 10 taxpayers in each group are in the lowest income bracket, 10 are in the next income bracket, and the remaining 10 are in the highest bracket. The amount of time (in minutes) needed to complete the returns is recorded and provided on the CSV File *IRS 3*. The data has the following columns.

Column 1: Group Number (1 = Low Income, 2 = Next Income Bracket, 3 = Highest Bracket)
Column 2: Times to complete Form 1 (first 10 rows = low income, next 10 rows = next income bracket, and last 10 rows = highest bracket)
Column 3: Times to complete Form 2 (same format as column 2)
Column 4: Times to complete Form 3 (same format as column 2)
Column 5: Times to complete Form 4 (same format as column 2)

a) How many factors are there? State them.

b) What are the levels for each factor? State them.

c) How many treatments are there?

d) At 5% significance level, is there an evidence of statistically significant interaction? Use the p-value from the ANOVA output and interaction plot to answer. If there is an interaction, describe it.

e) At 5% significance level, can we conclude that differences exist among the four forms?

f) At 5% significance level, can we conclude that taxpayers in different brackets require different amounts of time to complete their tax forms?

Question 4 *(18 points)*

Detergent manufacturers frequently make claims about the effectiveness of their products. A consumer-protection service decided to test the five best selling brands of detergent, where each manufacturer claims that its product produces the "whitest whites" in all water temperatures. The experiment was conducted in the following way. One hundred fifty white sheets were equally soiled. Thirty sheets were washed in each brand — l0 with cold water, 10 with warm water, and 10 with hot water. After washing, the "whiteness" scores for each sheet

were measured with laser equipment.

Column 1: Water temperature code
Column 2: Scores for detergent 1 (first 10 rows = cold water, middle 10 rows = warm, last 10 rows = hot)
Column 3: Scores for detergent 2 (same format as column 2)
Column 4: Scores for detergent 3 (same format as column 2)
Column 5: Scores for detergent 4 (same format as column 2)
Column 6: Scores for detergent 5 (same format as column 2)

The dataset is provided on the CSV file *Detergents*.

a) How many factors are there? State them.

b) What are the levels for each factor? State them.

c) How many treatments are there?

d) At 5% significance level, is there an evidence of statistically significant interaction? Use the p-value from the ANOVA output and interaction plot to answer. If there is an interaction, describe it in words.

e) Is there a reason to conduct the tests of the main effects?

f) Depending on your response to the previous part (above), conduct the tests of main effects if you need to. If you don't think you need to conduct the tests of main effects, then don't answer this part.