Statistics 411/511
Homework 5
Due Tuesday, November 3 by midnight Pacific time

- **Instructions:** Upload homework to Gradescope via Canvas (access specific homework assignments from the Assignments link on the Canvas course page). Your file must be a pdf document. Please see the end of the syllabus for formatting guidelines.
- The problems are assigned from the **third edition** of the textbook. If you have another edition, consult the copy on reserve at the library website for the homework problems.
- **Academic Integrity** You are encouraged to *discuss* the homework with other students, but what you turn in must be your own work in your own words. **DO NOT** copy someone else's homework. You may share ideas and R code, but do not share R output or written language. The syllabus contains details and links to OSU's Student Conduct Code and procedure for reporting suspected academic misconduct.
- **Please assign pages when submitting to Gradescope.** See Gradescope help for instructions and a video. Papers without assigned pages will lose 0.5 points.

1. Download the file Cotton.csv from Files>Homework on the course Canvas site. Read the data into RStudio (see the procedure in item 7 of Lab 2).

    This file contains data on five varieties of cotton fabric. Four 10-centimeter squares of each variety were randomly selected from a textile mill's inventory, and the luminance of each square was measured.

    The data file contains two variables:

    - Variety: A factor variable with five levels identifying the cotton variety
    - Luminance: A numeric variable with the luminance in candela per square meter (cd/m^2) for the fabric sample.

    (a) Make side-by-side boxplots of luminance scores for the five varieties. Include your R code and graph.

    (b) The ANOVA F-test is an extension of the two-sample t-test when there are more than two samples. The ANOVA F-test has the same three assumptions as the two-sample t-test. State each assumption and whether it is reasonable for the cotton data. Give a brief justification for each answer.

    (c) Display 3.5 on page 65 of the textbook illustrates the results of a simulation study that shows the two-sample t-test's results are reliable even if the ratio of variances is 1 to 4, provided the populations are normal and the sample sizes are equal. The ANOVA F-test is a generalization of the two-sample t-test, and an analogous results holds: provided the populations are close to normal, and the sample sizes are approximately equal, then the ANOVA F-test's results are reliable. Is an ANOVA F-test appropriate for the cotton data? Briefly explain.

    (d) State the null and alternative hypotheses to answer the research question, "does luminance vary among the five varieties?" Your hypotheses should be in terms of population parameters. Define any notation you use for the population parameters.

(Problem 1 continued on next page)

(e) Perform a one-way ANOVA to test your null hypothesis in (d). Include R code and the resulting ANOVA table.

(f) Write a statistical conclusion to report the results of your hypothesis test in part (e).

(g) Calculate a 95% confidence interval for the difference between the population mean luminance for the Giza68 and Giza69 cotton. Submit R code and output.

(h) Write a statistical conclusion for your confidence interval in (g)

(i) See pages 4 and 5 of Outline 1 where we first saw the idea of a scope of inference. Write a "scope of inference" for your conclusions in parts (f) and (h).