Assignment 10 — due April 22, 2019

A statistician, who had a large oven, conducted an experiment to see how different cooking practices affected the time it took to cook a pot roast in an oven. The factors studied were:

- Temperature: 325, 350, 375, or 400 degrees F.
- Type of pan: covered, uncovered, or with a rack inside.

The experiment was conducted over 3 months. On each Sunday when the statistician was not out of town, one of the temperatures was used. Altogether, 8 Sundays were used, and each temperature was used twice (2 days for each temperature). The assignment, to the days, of the temperatures was completely random. On each day, the statistician placed 3 roasts (of equal size) in the oven at the same time, such that there was one roast per pan, and the 3 pan types were used. The locations of the pans in the oven was completely random (and randomized separately on each day). The time (in minutes) that it took to cook each roast was determined by using a thermometer inside each roast. On the last Sunday, one data point is missing unfortunately, because the corresponding roast to be cooked had turned bad. The data are available on Canvas in file `potroast.csv`.

1. Plot the data. Show all variables in a single plot (using colors, point shapes, etc.). Based on this display, which factors seem most influential for cooking time? Does it seem that it was a good idea to cook all 3 pans together on each day?

2. Consider temperature as a categorical predictor for this question.

   (a) Which experimental design was used? Conduct an analysis of the data corresponding to this standard experimental design, including F tests for the effects of temperature and pan type. Compare the degrees of freedom in these F tests to the degrees of freedom expected from a balanced experiment with no missing data.

   (b) Compare the average cooking time at the 4 different temperatures. Explain why the t-tests are approximate.

3. Consider temperature as a numerical predictor for this question, with models where cooking time depends linearly on temperature.

   (a) Explain why, in this experiment, we might *not* want to allow for each day to have its own linear relationship with temperature.

   (b) Fit a model where each day is allowed to have its own intercept, for the linear relationship with temperature. Test for the presence of an interaction effect of temperature & pan type, 3 different ways:

      i. with a likelihood ratio test statistic and a chi-square null distribution,
      ii. with a likelihood ratio test statistic and parametric bootstrap to estimate the null distribution,
      iii. with an F test.

      Compare the three p-values with each other. Which one is the most accurate to report?

   (c) Using the same model as in (b), test for differences between days, 2 different ways:

      i. with a likelihood ratio test statistic and a chi-square null distribution,
      ii. with a likelihood ratio test statistic and parametric bootstrap to estimate the null distribution.

      Compare the three p-values with each other. Which one is the most accurate to report?

   (d) Fit a model as in (b) but with *no* interaction effect of temperature & pan type. Using this model, show how to use the output of the `summary` function (which might be the only output provided on the final exam) to

      i. report a confidence interval for the average cooking time change when the pan cover is removed;
      ii. predict the cooking time for a new roast cooked with the "Rack in pan" method, at 340F. Ignoring the uncertainty in this estimate, provide a rough

4. The goal of this question is to determine if the dependence of cooking time on oven temperature may be considered linear, as assumed in question 3. Consider two models for this question: the model from 3(d), and a model in which temperature is considered as a categorical predictor, like in question 2, but *without* an interaction effect between temperature & pan. Test the hypothesis of a linear trend with temperature, in 3 different ways:

   (a) with a likelihood ratio test statistic and a chi-square null distribution,

   (b) with a likelihood ratio test statistic and parametric bootstrap to estimate the null distribution,

   (c) with an F test.

   Again, compare the three p-values with each other. Which one is the most accurate to report?