

Problem set 6: More barley

S470/670

Upload your submission to Canvas by 11:59 pm, Friday 3rd December.

Note: There will be seven problem sets in all. Your lowest problem set score will be dropped when determining your grade.

In the Minnesota barley example we saw in class on 10/27, the major controversy was whether the Morris yields for 1931 and 1932 were mistakenly reversed. Cleveland asserts that the yields given for Morris 1931 were actually for Morris 1932, and vice versa. The file `minnesota.barley.yield.txt` contains a fuller set of information about the barley trials than is given in Cleveland, with data from 1927 to 1936:

- **yield:** barley crop yield in bushels per acre (a bushel is four pecks);
- **gen:** one of a number of types of barley;
- **year:** a year from 1927 to 1936;
- **site:** one of six locations in Minnesota.

With more data, we should be able to get a better idea of whether the Morris data can be explained as random variation or if it's a mistake.

Questions

1. Draw an appropriate faceted graph showing how barley yield varied by year at each location, using color as necessary. (Note: Lines might be better than dots here.) When looking at successive years, was it more common for the yields to move in the same direction at all locations, or was it more common for the yields to increase at some locations and decrease at others?
2. Fit a model with yield as the response and gen (variety), year, and site as explanatory variables, with the goal of determining whether Morris 1931-1932 is an anomaly. Justify why you chose this model and not some other one.
3. Draw plots of the fit and/or residuals with the goal of determining whether Morris 1931-1932 is a mistake, or whether it can be explained as natural variation. As best as you can tell, was there a mistake?