

Statistics 536 - Final Exam Case Study

1 Background

Tulips came to the Netherlands in the mid-16th century and quickly became a symbol of the region. To this day, Holland (a region along the west-cost of the Netherlands) is known as the “flower shop of the world.” Each year, around the end of April, hundreds of thousands of visitors from around the world flock to the region to participate in local tulip festivals bringing in millions of dollars of revenue for the region. Further, tulip exports account for approximately 25% of agricultural exports for the region.

Tulips are typically planted in the late Fall allowing for the bulb to go through a “chilling period” prior to blooming. Further, ideal growing conditions for tulips are in light, airy, well- drained soil. Once blooming begins, tulips require plenty of sunshine to reach full growth.

Changes to the Netherlands’ climate upsets these target growing conditions. The temperature is expected to rise at about twice the global average and precipitation is expected to increase. Further, with rising sea levels and the fact that 1/3 of the Netherlands is actually below sea level raises the risk of flooding to the area. In all, climate change threatens Netherlands tulip economy and thus we study how tulip growth may be affected.

Over the past several years, researchers have been studying various species of tulips to understand which species under which conditions are ideal for expected climate conditions moving forward. In one such experiment, 210 tulip bulbs of 12 different species (2510 bulbs total) were randomly split into 7 different treatment groups (30 bulbs per group per species). Each group was then assigned a chilling time of 0, 2, 4, ..., 12 weeks (in a lab refrigerator) after which the seeds were planted in separate pots in a greenhouse. The bulbs were then monitored to see if they germinated (bloomed) or not. While there are other details, we will assume that the laboratory conducted the experiment correctly.

2 Goals

Prepare a 20 minute oral presentation about your analysis (see the rubric for specifics). In this analysis, we want you to answer the following questions for the audience:

1. What is the effect of chilling time for the different species of tulips? Is it the same across the species?
Which species are the same/different?
2. Is there an ideal chilling time for each species? If so, is it the same for all species?
3. Given climate change conditions, winters are expected to decrease from 10 to 9 weeks in the coming few years. What effect will this decrease in chilling time have on the probability of germination for each species? Is it the same for all species?

3 Data File and Variables

Data are available in the file Germination.csv, and contains the following variables.

- Population - integer indicating one of 12 different species.
- YearCollected - year that the measurement was taken
- ChillingTime - the number of weeks that the specimen was chilled.
- Germinated - a binary variable indicating whether the specimen germinated (bloomed) or not.

4 Tips

- For your two proposed models, one should be from an additive/very-interpretable model (e.g. linear/logistic/GAM/etc.) and the other should be a very predictive model (machine learning). Both models should be appropriate for the data limitations and research questions though!
- Methods discussed in class this semester are sufficient. You don't need to find/try new methods.
- Audience: Me! Dr. Fisher. I hope this simplifies things and we won't need to act out parts.
- For presentations with slides, the abstract should just be bullet points instead of paragraphs. Actually bullet points should replace all paragraphs!
- Focus on the bold parts of the rubric:
 - EDA to find peculiarities/challenges in the data and in using the data to answer the research questions (with good plots/figures!).
 - Propose models that appropriately address both the data's limitations and the research questions.
 - Model Choice is based on IS/OOS performance and weighing interpretability vs. predictive power.
 - Answering research questions well, with good plots/figures when you can/when appropriate.
 - Did I mention using helpful visualizations? ;) As this is a presentation, use visualizations to your advantage! (and avoid using too much text on slides; no paragraphs).