

STAT-S632 – Midterm Exam – Take-home

Due March 9, 2018

Instructions:

Please read carefully the following instructions:

- Start your document with the statement: “With the exception of the class instructor, I have not had any form of communication about this exam with any other individual (including other students, teaching assistants, instructors, etc.)” and sign by hand (if work was done partially or completely by hand and scanned) or typing your name (if work was done all typed).
- You will be allowed to ask the instructor clarification questions about this exam only during class sessions. This ensures that all students receive the same information about the exam. No questions about this exam will be answered during individual office hours.
- You are allowed to use course resources without need of citations. The use of additional resources (e.g., papers, manuscripts, textbooks, internet, etc.) needs to be cited appropriately. Failure to cite resources could be penalized.
- Your document should be prepared in the following way:
 - Final answers should be included in the **first 2 pages** of your submission paper.
 - Supporting material (R syntax used or additional output, graphs, etc.) is also required and should be added after the first two pages. In addition, if any given output was used to answer questions, the answer should include a clear reference to the material location (e.g., figure number, table number, or at least page number). There is no limit of space for supporting material.
- Failure to follow any of the instructions above could result in reduction of your grades.
- Submit your solutions to Canvas as a single PDF file by 6pm on Friday, March 9th. Do not wait until the very last minute since late submissions will not be graded.

Questions

Using the data `takehome.txt`, researchers are interested in how different types of weeds are affected by the percentage of humidity and the number of days the lawn has not been treated with some type of weed-control. Researchers selected 45 sites in Bloomington. Randomly chosen sites were searched for different type of weeds. At each selected site, a $500ft^2$ area lawn was examined and the number of different types of weed (`Weed`) was determined. The key question is how the number of types of weeds is related to the percentage of humidity of the land (`Moist`) and the number of days of land untreated with weed-control products (`Days`).

Looking at the data, you could see that the percentage of humidity for each site is either 52% (or below) or 77% (or above). This is, potentially, information you might want to include in your model.

1. Using only the methods and techniques learned so far in S632 (and, methods learned in 631 if you consider this appropriate), propose an appropriate model. In your work, you should used as many as the methods and techniques you’ve learned so far that could be relevant.
2. Based on your chosen model, interpret at least 2 of the coefficients estimates.
3. Assume that your lawn is about $500ft^2$ in size with 83% humidity level and you have not used any weed-control produce for a whole year. Answer the following questions:
 - a. How many types of weeds would you expect to find?
 - b. What is the probability you are not going to observe any weeds?
 - c. What is the probability you may observe more than 3 types of weeds?
4. If your chosen model was obtained using `glm` in R, write the R code necessary to obtain the coefficient estimates without using `glm` and show that your results are equivalent to those using `glm`. If you did not use `glm` for your chosen model, write the R code necessary to replicate the results of `m1 = glm(Worms ~ Days*Moist, family = "poisson", data = takehome)`, where `takehome` is the data frame for `takehome.txt`. (Hint: It might be easier just to modify some the code provided in class)