STAT 515 – Spring 2018 – Take-home final exam

Note: For this final exam, **you are not allowed to receive help from *anyone except me* on the exams.** For example, you may not talk to other students about the exam problems, and you may not look at other students' exams. **Violations of this policy may result in a 0 on the exam, an F for the course, and/or punishment by the USC Office of Academic Integrity.**

Problems 1-5 below involve answering some statistical questions about data and writing mini-reports detailing your conclusions. Some (but not all) of the problems can be made easier by the use of SAS. You can run the SAS code given at the following website to input the data into SAS data sets:
`http://people.stat.sc.edu/hitchcock/FinalExamSAScodeSTAT515Spring2018.txt`

I have not provided all the SAS code that may be helpful to do an appropriate analysis, because (for **some** problems) I would like you to choose the method of analysis that you believe is appropriate and you may write your own small bit of code to perform this analysis (the code can be based on example code we have used in class examples, homework example, etc., and applied to these exam data sets). Not all the problems require you to use SAS, but for certain problems it would be very helpful.

Your mini-reports should be typed in paragraph form and may include relevant graphs where appropriate. For each mini-report, the amount of actual text (not counting plots and graphs) certainly does not need to be more than one page in length. Some questions can be suitably answered with mini-reports substantially shorter than one page. Please do not include any computer code within your mini-reports. If you wish to include computer code, please put it in an appendix at the end of your document. If you want to include details of calculations (which is your choice … it may or may not be helpful), it may be good to put those in an appendix as well rather than cluttering up the mini-reports.

Your reports will be graded partly on the quality of the statistical analysis that you do, but **just as importantly** on your ability to communicate your conclusions **in your own words** clearly and concisely. Specifically, each problem will be worth 10 points, for a total of 50 points. Each mini-report will be graded based on the writing/communication (out of 5 points) and the appropriateness and the correctness of the statistical work (out of 5 points).

Note that some of these problems are somewhat open-ended and there may be more than one way to appropriately answer the question. In such cases, often it is the way you justify your method and explain your findings in your mini-report that is the key to a great answer. Also recall that another aspect of a great answer is checking whether your chosen approach is valid for the data set and that the data meets the assumptions of the method.

Note that part of what I am testing for on in this take-home exam is the ability to pick an appropriate method to answer a statistical question. So I can clarify the phrasing of a question, but if you ask me to tell you what method to use on a problem, I am likely to decline to answer that. Similarly, I will not tell you what SAS code to write in order to do something (I expect you to know that based on the examples we have done). However, if your SAS code is producing errors, I **may** be able to help you with the code if it is a **small** error. If your code is nowhere near correct, I will likely just refer you to look at the SAS examples on the course website. In any case, please start early on the exam, so that if any clarification is needed, you can ask **early** and not at the last minute! I may not always be able to reply right away, so ask early!

NOTE: For all problems requiring inference, you can use $\alpha = 0.05$ or $1 - \alpha = 0.95$ unless otherwise stated.

Problem 1: A math instructor is comparing two methods of teaching, a modern method and a traditional method. She uses the modern method for one class (of size 25 students) and the traditional method for another class (of size 25 students, who are different people from the first class). The two sets of students can be considered samples from the two populations of students that would take this class taught with the two methods. The question of interest is to compare the population mean final exam scores for the two methods. The instructor does not know ahead of time which method is likely to produce better results. The data and some SAS code are given on the course website. Note from the output that the sample mean of the "modern method" scores is 77.776, with a sample standard deviation of 9.298. Also, the sample mean of the "traditional method" scores is 81.972, with a sample standard deviation of 1.976. Write a mini-report describing a data analysis of your choice and your conclusions to answer the instructor's research question.

Problem 2: A scientist is interested in studying the weight of the larva of the Hercules beetle, found in Central and South America. The scientist takes a trip to the rainforest and gathers a sample of 22 specimens of Hercules beetle larvae. The scientist would like to obtain point and interval estimates of both the population mean weight and population standard deviation, for the population of Hercules beetle larvae. The data and some SAS code are given on the course website. Note that the sample mean is 114.727 grams and the sample standard deviation is 6.263. Write a mini-report describing a data analysis of your choice and your conclusions to answer the scientist's research questions.

Problem 3: A researcher wants the study the resting heart rates of three groups: nonsmokers, light smokers, and heavy smokers. He takes a sample of 14 nonsmokers, a sample of 14 light smokers, and a sample of 14 heavy smokers. Of interest is whether the mean resting heart rates are the same across the three populations. If the mean rates are different, the researcher would like some insight into which populations have different mean rates and the likely ordering of the population means. The data and some SAS code are given on the course website. Note that the sample mean heart rate for the nonsmokers was 60.193, with a sample standard deviation of 7.0005; the sample mean heart rate for the light smokers was 72.907, with a sample standard deviation of 5.263; the sample mean heart rate for the heavy smokers was 77.457, with a sample standard deviation of 6.269. Write a mini-report describing a data analysis of your choice and your conclusions to answer the research question.

Problem 4: An admissions officer at a Midwestern college is investigating the relationship between ACT math test score and college GPA. She is hoping to develop a model to predict college GPA using math ACT score. She has obtained a sample of 30 current students whose college GPA and math ACT scores are known. The data and some SAS code are given on the course website. The researcher would like to know whether a relationship exists between these two variables, what is the nature of the relationship, and whether math ACT is a useful predictor of college GPA. In particular, she knows of a potential admittee with a math ACT score of 23 whose GPA she wants to predict with a point prediction and a prediction interval. Develop a detailed analysis of these data addressing the officer's question and any other aspects of the data that you find relevant to the model. Write a mini-report describing your data analysis of choice and your conclusions to answer the research questions.

Problem 5: A study examined a potential relationship between the categorical variables: survival of coronary heart disease (CHD) patients, and pet ownership. The population of interest was all patients with CHD. A random sample of CHD patients was taken and each patient was classified according to (1) whether he or she survived for one year after diagnosis, and (2) whether he or she owned a pet. Of the patient who did not own a pet, 56 survived for at least one year, and 22 died within one year. Of the patients who owned a pet, 96 survived at least one year, and 16 died within one year. The question of interest is whether pet ownership is associated with survival, or whether they are independent. The researchers were also interested in learning whether pet ownership might actually improve survival

chances. Write a mini-report describing a data analysis of your choice and your conclusions to answer the research questions.

The final exam will be due by Tuesday, May 8, by 4:00 p.m. You should send me a single document (a Word document or pdf file is fine) with your mini-reports via email before the deadline.