STAT 516 – Spring 2020 – Take-home midterm exam

Note: For this 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 1-4 below involve answering some statistical questions about data and writing mini-reports detailing your conclusions. To answer these questions best, you should use SAS to help analyze the data. 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/MidtermExamSAScodeSTAT516Spring2020.txt`

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

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. The average mini-report might be two or three paragraphs in length. Please do not include any computer code within your mini-reports! If you wish to include computer code (it is not necessary to do so), please put it in an appendix at the end of your document. If you want to include any details of calculations (which is your choice ... it may or may not be helpful), then you should probably put those in an appendix as well rather than cluttering up the mini-reports. Your mini-reports can certainly include numerical justification (such as relevant summary statistics, test statistic values and/or P-values) of your conclusions, but not details of any intermediate calculations you might do.

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 40 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 set meets the assumptions of the method (and if it doesn't, then making appropriate adjustments)!

Note that part of what I am testing for on 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 producer of medical masks is doing an experiment to investigate the effect that the material of the mask and the thickness of the mask have on protection against the COVID-19 virus (as measured by percent of 0.3-micron particles blocked; a higher percentage is better). The three material types examined are cotton, polyester, and polypropylene. The three thickness levels examined are 0.25 mm, 0.40 mm, and 0.55 mm. The researcher also wants to know whether the material type and the thickness interact significantly. If the factor(s) significantly affect protection, then we want to know which levels of the factors differ significantly (or which factor level combinations, if there is significant interaction). The data and some SAS code are given on the course website. Write a mini-report describing a data analysis of your choice and your conclusions to answer the producer's research questions.

Problem 2: A CEO of a retailer that has many stores around the country is investigating which factors affect sales (in thousands of dollars). The factors considered are store location (urban, suburban, rural) and type of advertising strategy used (none, TV only, radio only, TV and radio). The CEO wants to know whether these factors affect sales, and also whether the effect of advertising depends on the store location. Another critical issue is: TV advertisements are very expensive in urban areas. The CEO wants to formally compare (specifically in urban areas) the strategies that include TV ads to the strategies that do not include TV ads, in terms of their mean sales. Data were gathered from a total of 24 stores following an experiment that used different advertising strategies in the various areas. The data and some SAS code are given on the course website. Write a mini-report describing a data analysis of your choice and your conclusions to answer the CEO's research questions, and your recommendations to the CEO.

Problem 3: A biostatistician is conducting a clinical trial to compare five drugs (labeled A, B, C, D, and E) in terms of their mean systolic blood pressure reduction. We wish to know whether there are differences among the five drugs in terms of how much they reduce blood pressure on average. In particular, note that drugs A and B are "existing" drugs and drugs C, D, and E are "new" drugs, so we want to formally assess whether the new drugs provide greater blood pressure reduction, on average, than the existing drugs. The experiment was conducted on a total of 30 subjects. The data and some SAS code are given on the course website. Write a mini-report describing a data analysis of your choice and your conclusions to answer the research questions.

Problem 4: An agricultural researcher is investigating the relationship between fertilizer amount and yield of a corn crop, under two different growing conditions. The levels of fertilizer amount (in pounds) are: 40, 60, 80, and 100. The two growing conditions are "sunny" and "shaded". We want to know whether the effect of fertilizer amount on mean yield is different under the two growing conditions. If so, then which pairs of factor level combinations are NOT significantly different? If not, then which levels of fertilizers are significantly different? [Also, if it is of interest for the researcher to predict the corn yield if 90 pounds of fertilizer were used in shaded conditions, what approach would you take, using these data? Do you make any extra assumption to answer this question?] The data and some SAS code are given on the course website. Write a mini-report describing your data analysis of choice and your conclusions to answer the research questions, and giving your recommendation for the best agricultural practice to get good corn yields.

This exam will be due by Wednesday, April 1, by 11:59:00 p.m. **You should upload a single document** (a Word document or pdf file is fine) with your mini-reports via Blackboard before the deadline.