UNM Statistics Qualifying Exam Take-Home
January 2019

Due 12:00pm Jan 11 Friday, 2019. Return to Ana Parra Lombard in the Math/Stat Dept Office SMLC 395.

Directions: The answer to each problem should be presented as a summary. It should be word processed and double spaced, and be identified by your "Code Word" (do not include your UNM ID and name). **A suggested length of the report to each problem is no longer than 3 pages**. Create brief, well-organized appendixes (not included in 3 pages) for each problem.

In your data analysis, RAW AND UNINTERPRETED COMPUTER OUTPUT IS UNACCEPTABLE. You should have a caption by every figure and table that describes it and tells the reader briefly what you see. Organize the sections to tell the story you uncovered, not the circuitous path you may have taken to get there. Remember that even that best data analysis is worthless if your reader cannot understand it.

You may **not** consult any other person when working on this exam or discuss your exam with anyone else regardless of whether or not the person is taking the exam. You may use your course notes as well as any available books or web resources for the exam. Questions pertaining to clarification about these questions can be directed to Guoyi Zhang, gzhang12@math.unm.edu.

Problem 1. (50pts) The data for this problem is available at `http://www.math.unm.edu/~gzhang12/data/psa.txt`

Serum prostate-specific antigen (PSA) was determined in 97 men with advanced prostate cancer. PSA is a well-known screening test for prostate cancer; the goal of this study is to examine the correlation between level of PSA and a number of clinical measures for men who were about to undergo radical prostatectomy. Following variable PSA are the measures: cancer volume, prostate weight, patient age, the amount of benign prostatic hyperplasia, seminal vesicle invasion, capsular penetration, and gleason score.

Using the data, build an appropriate regression model to analyze how strongly the measures given above correlate with PSA. There are 97 men (subjects) in this study. You are expected to select a random sample of 70 men to be used in model-building. Summarize your findings. Make sure you tell the reader what you found and how you found it. Make sure you include important results in your report and put some necessary outputs in appendix. Provide PSA prediction for the remaining 27 subjects and compare it to the true values.

Problem 2. (50pts) The data for this problem is available at `http://www.math.unm.edu/~gzhang12/data/prison.txt`.

Subjects from the general population of Central Prison, Raleigh, NC, volunteered for an experiment involving an "isolation" experience. (Those convicted of felon offenses are housed in this facility.) The experimental treatment exposed inmates to combined sensory restriction and suggestion. The intent was to reduce the psychopathic deviant T scores (Pd T), Scale 4 of the Minnesota Multiphasic Personality Inventory (MMPI) test.

Briefly, the three treatments consisted of

(a) Four hours of sensory restriction plus a 15 minute "therapeutic" tape advising that professional help is available.
(b) Four hours of sensory restriction plus a 15 minute "emotionally neutral" tape on training hunting dogs.
(c) Four hours of sensory restriction but no taped message.

Forty-two subjects were assigned to one of the three treatment groups (for a total of 14 in each treatment group). For each subject the MMPI was administered before and after the experimental treatment. Pre-treatment and post-treatment values of Pd T scores are given below for the 42 individuals.

Analyze the data provided by this experiment. In addition to analyses and comments arising from your own curiousity, you must address the following as part of your write-up. It is recommended that you structure your write-up similar to the order of the questions below.

(a) (6 pts) What statistical design is being used, and why?
(b) (1 pts) Is there blocking? If so, what is/are the block(s)?
(c) (1 pts) What is/are the nuisance factor(s) to be "averaged out" in the design?
(d) (1 pts) What is/are the treatment(s)?
(e) (1 pts) What is/are the outcome(s)/response(s)?
(f) (5 pts) Plot the data (not only summaries of the data, such as only means) in a way that helps you understand what the effects are.
(g) (5 pts)
Write out the best full statistical model (in notation, defining the notation you use) and state the model assumptions.
(h) (10 pts) Fit the model written in the previous part (that is, effects in fitted model should be as in the model specification above), and assess and address deviations from model assumptions. This may be an iterative process. Summarize each model fit and the evidence for the decisions made to arrive at your final model, consider moving intermediate model fit details to the appendix. (Note: If model assumptions are not met, try to address that. If you can not address unsatisfied model assumptions, mention this and continue as though the model assumptions are met.)
(i) (5 pts) State and conduct statistical tests for the parameters, and interpret the test results.
(j) (2 pts) How many degrees-of-freedom are allocated to each source of variation in the final model?
(k) (5 pts) Interpret at least one interesting parameter value.
(l) (8 pts) Perform pairwise comparisons based on your final model and summarize which pairs of treatment combinations are different.