STATISTICS MASTERS/Ph.D.-QUALIFYING EXAM: TAKE HOME
August, 2019

Directions: The answer to each problem should be presented as a summary. It must be word processed, double spaced, and identified with your "Code Word" in the header on each page; *do not include your UNM ID or name.* The report for each problem should be no longer than 3 pages. You may create brief, **well-organized** appendixes for each problem. Remember, the appendix is not the report and *will only be examined* if the report draws interest to it and material is easy to find. In your data analysis, raw uninterpreted computer output will be graded as the dross it is. You should have a caption for every figure and table; one that describes it and briefly tells the reader why it is of value. Organize your sections to justify the validity of what you uncovered and the methods you used to uncover it. We want a summary of what you think is important, not a diary of how you spent your time. Remember that even the best data analysis is worthless if your reader does not understand it, so you are being graded on presentation as well as statistical content. 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. (No matter what you think, you will not find *these* problems on the web.) Questions can be directed to Ronald Christensen, fletcher@stat.unm.edu. Email solutions by **3 PM, Mon Aug 19, 2019** to Ana Parra Lombard, aparra@math.unm.edu, Department of Mathematics and Statistics, University of New Mexico. Please do not turn in a physical copy of your solutions.

1. Your goal is to predict the body *Density* determined from underwater weighing of 250 men based on their

   - *Age* (years),
   - *Weight* (lbs),
   - *Height* (inches),
   - *Neck* circumference (cm),
   - *Chest* circumference (cm),
   - *Abdomen* circumference (cm),
   - *Hip* circumference (cm),
   - *Thigh* circumference (cm),
   - *Knee* circumference (cm),
   - *Ankle* circumference (cm),
   - *Biceps* (extended) circumference (cm),
   - *Forearm* circumference (cm), and
   - *Wrist* circumference (cm).

   Find the best predictive statistical model that you can, justify the appropriateness of your model, and discuss what conclusions can be reached from this model.

   The data are in comma-separated values (csv) format.
   https://math.unm.edu/sites/default/files/files/qual-exams/Statistics/stat_qual_201908_takehome_dat1_density.csv

2. An experiment has been designed to study a measure of plant growth (y) in response to three (3) soil treatments ($C = 1, 2, 3$). Unfortunately, the three (3) greenhouses available ($A = 1, 2, 3$) have different soil and crop histories, which may have an effect on the response and may interact with the soil treatments. Nine (9) areas ("Block") are available throughout in each greenhouse in which to conduct a set of three soil treatments. Each block is organized into a space of roughly 9 square meters, where treatments are randomized into 1-by-3 meter lengths and partitioned to avoid treatment cross-contamination.

   The data are in comma-separated values (csv) format.
   https://math.unm.edu/sites/default/files/files/qual-exams/Statistics/stat_qual_201908_takehome_dat2_growth.csv

   Analyze the data provided by this experiment. In addition to analyses and comments arising from your own curiousity, you must address the following items as part of your write-up.

   (a) What statistical design is being used, and why?

   (b) Plot the data (not only the means, but every data point on well-labelled axes) in a way that helps to understand what the effects are.

   (c) Write out the best full statistical model (in notation, defining the notation you use) and state the model assumptions.

   (d) Fit the model written in the previous part (that is, estimates in the fitted model should correspond to parameters in the model above), and assess and address deviations from model assumptions. This may be an iterative process. Summarize each model fit and discuss the evidence guiding decisions to arrive at your final model. (Note: If model assumptions are not met, try to address those. If you can not address unsatisfied model assumptions, admit to this and continue as though the model assumptions are met.)

   (e) State and conduct statistical tests for the parameters, and interpret the test results.

   (f) Perform pairwise post-hoc comparison tests based on your final model and summarize which pairs of treatment combinations are different. If there is a meaningful interaction, conditional comparisons may be appropriate.

   (g) What do you recommend to the farmer who wishes to maximize growth?