STATISTICS MASTERS/Ph.D.-QUALIFYING EXAM: TAKE HOME
January, 2023

General directions

Complete both problems in this exam. Your report is to be typed, double spaced, no smaller than ten-point font with one-inch margins, and should be identified by your "CODE WORD" in the header on each page; *do not include your name or UNM ID number.* Each problem is to be no longer than four pages, and an additional four-page appendix is allowed for each problem but will be examined only at the discretion of the graders; the better constructed your appendix is with cross-references from the text, the more likely it is to get examined. In your data analysis, raw uninterpreted computer output will be graded as the dross it is.

Write your answers completely, but concisely. Insert tables and figures to support your points. Tables and figures should be well-labeled and cross-referenced from text, such as, "in Table 1 …", or if in the appendix, "in Table A1 …", and each should have a caption that describes it and briefly tells the reader why it is of value. Figures should include appropriate symbols suitable for black-and-white reproduction (that is, avoid use of color if possible; consider symbols, line types, and distinct shades of gray to distinguish categories or values).

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.

As necessary:

1. Plot and describe the data (that is, plot all the individual observations, in addition to summaries of data you might present with the results, such as means and confidence intervals).

2. Clearly define population parameters and sample statistics.

3. Clearly specify hypotheses tested and explicitly state the associated model at least once (i.e., write the model equation).

4. Define and assess method assumptions.

5. Write a coherent evidence-based conclusion that a layperson can understand.

You may use your course notes as well as any available books or web resources on general statistical methods for the exam. 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 nor are you allowed to use the internet to find analyses of these data.

Any points of clarification can be directed to Prof. Erik Erhardt, `erike@stat.unm.edu`.

Email solutions by **3 PM, Fri Jan 13, 2022** to Ana Parra Lombard, `alombard@unm.edu`, Department of Mathematics and Statistics, University of New Mexico. Please do not turn in a physical copy of your solutions.

Problem 1, Grain

In the spring of the year, a randomized complete block design was set up to investigate yields of three grain varieties (A, B, C). Six blocks of land were obtained for the experiment and each was divided into three plots and the varieties were randomly assigned to the three plots. Later in the year it was decided to look at the effects of four different cutting dates on yields (D1, D2, D3, D4), so each plot was subdivided into 4 sections and these sections were randomly assigned one of the four different cutting dates. Identify the appropriate experimental design. Analyze the data. Discuss which interactions, if any, are of use in analyzing these data. Are there outliers? Are transformations necessary? What would you recommend to the farmer?

Download the data from
https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202301_stat_qual-takehome_dat1.csv.

Problem 2, Sleep

The USA BLS.gov American Time Use Survey (ATUS) measures the amount of time people spend doing various activities, such as paid work, childcare, volunteering, and socializing. We use a small subset of a version of the data summarized by Prof. Erhardt. Using the data, build an appropriate explanatory regression model (ignoring sampling-based weighing) to understand the associations between the number of hours slept and the covariates provided. Restrict your investigation to those who went to sleep and those who work and earn an hourly wage. Treat the education variable as numeric by converting the categories to the number of years educated. Recode values to missing when appropriate. Please understand the data and make sensible data coding decisions prior to your analysis — do not blindly start analyzing the data. Make sure that you carefully assess all model assumptions.

Write a succinct, coherent, and complete summary of your analysis, including a coherent interpretation of model effects.

Finally, also predict the expected hours of sleep of these two people:

1. a 20-year-old Metropolitan Female with a Master's degree who earns 2000 per hour with 1 child who spends 30 minutes alone and 180 minutes with family members who works 40 hours at a main job with no other job, and

2. a 70-year-old Non-metropolitan Male who graduated from High school who earns 1200 per hour with 0 children who spends 240 minutes alone and 0 minutes with family members who works 20 hours at a main job and 10 hours at another job.

Data:
TUCASEID : ATUS Case ID (14-digit identifier)
t0101 : Sleeping (minutes)
TESEX : Sex
TEAGE : Age
GTMETSTA : Metropolitan status (2000 or 2010 definitions)
PEEDUCA : What is the highest level of school you have completed or the highest degree you have received?
TRERNHLY : Hourly earnings (2 implied decimals, thus 2000 is $20.00)
TRHHCHILD : Presence of household children < 18
TRTALONE : Total time respondent spent alone (in minutes)
TRTHHFAMILY : Total time respondent spent with household family members (in minutes)
TEHRUSL1 : How many hours per week do you usually work at your main job?
TEHRUSL2 : How many hours per week do you usually work at your other job(s)?

Download the data from