

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

August, 2022

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 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 cross it is.

Write your answers completely, but concisely. Insert tables and figures to support your points. Tables and figures should be well-labelled 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 the mean 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 Aug 19, 2022** 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.

Problem 1, Anchovies

The estimated numbers of 4 Types of anchovies (A, B, C, D) caught in 12 successive Days (hauls) in a midwater fishing experiment. Analyze these data, focusing attention on potential differences in catches across days, and across plankton types. Be complete.

Please address the following questions as part of your write-up:

1. What statistical design is being used, and why? Could a better design have been used, and why or why not?
2. Is there blocking? If so, what is/are the block(s)?
3. What is/are the nuisance factor(s) to be "averaged out" in the design?
4. What is/are the treatment(s)?
5. What is/are the outcome(s)/response(s)?
6. Plot the data (not only summaries of the data, but all of the values) in a way that helps you understand what the effects are.
7. Write out the statistical model (in notation) and fit the model parameters.
8. How many degrees-of-freedom are allocated to each source of variation?
9. State and assess model assumptions. (If 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.)
10. State and conduct statistical tests for the parameters, and interpret the test results.
11. Discuss anything else of interest and address the original goal of the experiment.

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202208_stat_qual-takehome_dat1.csv.

Problem 2, Oxycodone

We would like you to construct a regression model for per capita oxycodone consumption by state (excluding AK and HI). Your job is to determine what, if anything, seems to drive consumption.

The variables in the data follow.

index : list of entries from 1 to 96

region : South (S), Mountain west (M), Midwest (MW), West Coast (WC), Northeast (NE), or Great Plains (GP).

year : only two.

cpi : consumer price index

population : Population of the state

oxy : average number of pills consumed per capita times 10.

income : total personal income for state.

tax : Average tax excluding sales taxes.

price : Average price (per prescription of 12 pills in dollars times 5) during the year including sales tax.

taxs : Average tax including sales taxes.

At least one of the faculty expects you to cover **at a minimum** all the items in <https://www.stat.unm.edu/~fletcher/SUPER/ALGOR.pdf>. (This was written several years ago for his classes, *not for this exam*.) In particular there are quite a few obvious relationships between various predictor variables that make them less interesting than they could be.

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202208_stat_qual-takehome_dat2.csv.