

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

January, 2025

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 dross 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. Avoid restating the problem in your introduction; spend more space on statistical issues and evidence-based decisions.
2. 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).
3. Clearly define population parameters and sample statistics.
4. Clearly specify hypotheses tested and explicitly state the associated model at least once (i.e., write the model equation). The model equation should be in the same parameterization as the results you present (that is, it will probably have a baseline intercept instead of a grand mean).
5. Define and assess method assumptions.
6. 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 as a pdf file by **3 PM, Fri Jan 17, 2025** 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, Student-Teacher Ratios

The UNESCO Institute of Statistics collects country-level data on the number of teachers, teacher-to-student ratios, and related figures.

Reducing class size has been a long-debated strategy to improve student achievement, with mixed research results. While some studies show no clear link, most find that class size reduction in primary grades (K-3) lead to improved student performance.

Using the data provided, compare the average student ratio by education level (indicator) adjusting for Country. Only use data for the year with the most observations and exclude the country "World". Control for multiple comparisons using a 95%-level Tukey's HSD.

Write a succinct, coherent, and complete summary of your analysis that addresses the goal. Remember to follow the "General directions".

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202501_stat_qual-takehome_dat1.csv.

Problem 2, Albuquerque Ozone

Historical Ozone Monitoring Data from the New Mexico Environment Department was collected for Albuquerque International Airport, NM, daily for three years: 7/1/21 – 6/30/24. Additionally, temperature and precipitation data from the National Centers for Environmental Information, National Oceanic and Atmospheric Administration, was collected for Bernalillo County, NM, for the same date range. The two datasets were joined by date and specific columns were selected.

1. **DATE:** Date
2. **AWND:** Average daily wind speed (mph)
3. **PRCP:** Precipitation (inches)
4. **SNOW:** Snowfall (inches)
5. **SNWD:** Snow depth (inches)
6. **TAVG:** Temperature, average of hourly values (F)
7. **TMAX:** Maximum temperature (F)
8. **TMIN:** Minimum temperature (F)
9. **2ZJ Bernalillo:** Ozone (O3) concentration (ppm)
10. **DaysJuly1:** Days since July 1 (as an example of a calculation using Date)

Treating observations independently, develop an explanatory regression model for Ozone concentration conditional on weather and seasonal measurements. Interpret each of your model effects clearly. Remember to follow the "General directions".

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202501_stat_qual-takehome_dat2.csv.