

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

August, 2021

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 dress 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 model/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. Guoyi Zhang, gzhang12@math.unm.edu.

Email solutions by **3 PM, Tue Aug 17, 2021** 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: Sleep

How does the average amount of Sleep a person gets depend on Region of the country, Metro area, Sex, and Age?

The American Time Use Survey (ATUS) is a time-use survey of Americans, which is sponsored by the Bureau of Labor Statistics (BLS) and conducted by the US Census Bureau. Respondents of the survey are asked to keep a diary for one day carefully recording the amount of time they spend on various activities including working, leisure, childcare, and household activities. The survey has been conducted every year since 2003.

Included in the data are main demographic variables such as respondents' age, sex, race, marital status, and education. The data also includes detailed income and employment information for each respondent. While there are some slight changes to the survey each year, the main questions asked stay the same. We are using data for 2019 and you can find the data dictionaries for each year on <https://www.bls.gov/tus/dictionaries.htm>.

The data have been joined and formatted for analysis for this exam. The “ATUS 2019 Activity summary file” was joined with “ATUS-CPS 2019 file” and all activities within a category were added together, such as all variables starting with “t01” into a new variable “act01”.

Variables in this analysis include:

- Region = “Northeast”, “Midwest”, “South”, “West”
- Metro = “Metropolitan”, “Non-metropolitan” (“Not identified” were combined with “Non-met”)
- Sex = “Male”, “Female”
- Age = “15-29”, “30-49”, “50-99” (summarized from integer ages)

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202108_stat_qual-takehome_dat1.csv.

Problem 2: Fuel

The data were collected by Christopher Bingham for the American Almanac of 1974 with the purpose to study fuel consumption. The data set contains 48 observations with 10 variables.

The variables are described as follows:

- V1: STATE
- V2: AREA (sq. mi.)
- V3: LAND (sq. mi.)
- V4: 1971 population in thousands [POP]
- V5: 1972 motor fuel tax rate, in cents per gallon [TAX]
- V6: 1971 thousands of licensed drivers [NLIC]
- V7: 1972 per capita income in thousands of dollars [INC]
- V8: 1971 thousands of miles of federal-aid primary highways [ROAD]
- V9: proportion of population with driver's licences [DRIVERS = $100 * (NLIC / POP)$]
- V10: gallons of fuel consumed per person [FUEL]

Using the data, build an appropriate regression model to predict FUEL using a subset of the rest variables, making sure that you carefully assess all assumptions. Write a succinct, coherent, and complete summary of your analysis.

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202108_stat_qual-takehome_dat2.csv.