

UNM Statistics Qualifying Exam
Due: 3 P.M., Mon Jan 16, 2017

Jan 2017

Name: _____
CODENAME

Qual Take Home (100 points) 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 CODENAME (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.

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 ...”. 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). Computer output without explanation will not be reviewed. 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 **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. If including computer text tables where alignment is important, then please use a fixed-width font, such as `Courier`, for that text. Any points of clarification can be directed to Prof. Erik Erhardt, `erike@stat.unm.edu`.

Due: Email Ana Parra Lombard <`aparra@math.unm.edu`> with solutions by 3 P.M., Mon Jan 16, 2017, Department of Mathematics and Statistics, University of New Mexico. Please do not turn in a physical copy of your solutions.

(50^{pts}) **1. Home prices**

The dataset for this problem is on median property prices in 2008 and possible covariates for 269 Intermediate Geographies (IG) that make up the Greater Glasgow and Clyde health board in Scotland.

The variables in the dataset are *price*, *crime*, *rooms*, *sales*, *driveshop*, and *type*, where *price* is median property price in each IG, *crime* is the number of crimes per 10,000 people in each IG, *rooms* is the median number of rooms in a property in each IG, *sales* is the percentage of properties that sold in each IG in a year, *driveshop* is the average time taken to drive to a shopping center in hours, and *type* is the predominant property type in each IG with levels: detached, flat, semi, and terrace.

Data: www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201701_pr1-DATA_pricedata.txt

- (a) (20 pts) Use a model selection technique to select a model to predict *price* using variables among *crime*, *rooms*, *sales*, *driveshop*, and *type*. Discuss which variables should be retained and which should be dropped and whether there is evidence for any interactions between variables.
- (b) (10 pts) Assess deviations from model assumptions; if the assumptions are violated, try to address those concerns, rebuild the model and reassess the assumptions.
- (c) (5 pts) Write out the statistical model (in notation) with all parameters retained in the final model. Describe in words of each covariate's effect.
- (d) (10 pts) Perform hypothesis tests of the main effects in your final model. Make sure to write down the null and alternative hypotheses, the test statistics, and the conclusions.
- (e) (5 pts) Use your final model to give an estimate of the mean price (or a function of the mean price if you have done transformation of the response variable in your final model) for an IG that has 556 crimes per 10,000 people, a median of 3 rooms in a property, 70% of properties being sold in a year, 1.5 hours taken to drive to a shopping center on average, and with flat as the predominant property type. Give a 95% confidence interval for this estimate.

(50^{pts}) 2. Horsepower test

A comparison of Fuel type (1, 2, 3) on engine Power is performed using five small Engines (A, B, C, D, E).

Data: `www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201701_pr2-DATA_horsepower.csv`

Fuel	A	B	C	D	E
1	0.155	0.163	0.159	0.202	0.105
2	0.166	0.172	0.170	0.215	0.138
3	0.155	0.162	0.163	0.189	0.127

Analyze the data provided by this experiment. In addition to analyses and comments arising from your own curiosity, please address the following as part of your write-up. It is recommended that you structure your write-up similar to the order of the questions below.

- (6 pts) What statistical design is being used, and why?
- (1 pt) Is there blocking? If so, what is/are the block(s)?
- (1 pt) What is/are the nuisance factor(s) to be “averaged out” in the design?
- (1 pt) What is/are the treatment(s)?
- (1 pt) What is/are the outcome(s)/response(s)?
- (5 pts) Plot the data (not only summaries of the data) in a way that helps you understand what the effects are.
- (5 pts) Write out the best full statistical model (in notation, defining the notation you use) and state the model assumptions.
- (10 pts) Fit the model written in the previous part (that is, effects in fitted model should be as in the model specification above), and assess and address deviations from model assumptions. This may be an iterative process. Summarize each model fit and the evidence for the decisions made to arrive at your final model, consider moving intermediate model fit details to the appendix. (Note: If model 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.)
- (5 pts) State and conduct statistical tests for the parameters, and interpret the test results.
- (2 pts) How many degrees-of-freedom are allocated to each source of variation?
- (8 pts) Perform pairwise comparisons based on your final model and summarize which pairs of treatment combinations are different.
- (5 pts) Discuss anything else of interest, and address the original goal of the experiment.