UNM Statistics Qualifying Exam Take-Home
Jan 2016

Due 4:00pm Jan 13 Wednesday, 2015. Return to Ana Parra Lombard in the Math/Stat Dept Office SMLC 395.

Directions: The answer to each problem should be presented as a summary. It should be word processed and double spaced, and be identified by your "Code Word" (do not include your UNM ID and name). **A suggested length of the report to each problem is no longer than 3 pages**. Create brief, well-organized appendixes for each problem.

In your data analyses, RAW AND UNINTERPRETED COMPUTER OUTPUT IS UNACCEPTABLE. You should have a caption by every figure and table that describes it and tells the reader briefly what you see. Organize the sections to tell the story you uncovered, not the circuitous path you may have taken to get there. Remember that even that best data analysis is worthless if your reader cannot understand it.

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. Questions pertaining to clarification about these questions can be directed to Yan Lu, luyan@math.unm.edu.

1. The data set `chile_takehome.txt` `www.math.unm.edu/~james/chile_takehome.txt` gives data on chile harvested in New Mexico in 2011 and 2013 (data for 2012 is unavailable). The variables in the data set are Year, Variety (usually where the chile was harvested, such as the name of a pueblo), Replicate for different plots (there were typically different plots within each location), a Fruit ID within each plot, Length of the chile pod in cm, Width in cm, Locule (number of chambers in the chile pod), and wall Thickness in mm for the chile pod. The chile varieties are Alcalde (coded as 1), Chimayo (2), and Casado (3).

   Answer the following questions.

   (a) Make a plot of Width versus Length, with a separate subfigure for each number of locales using the 2011 data. Using plot symbols to distinguish the year and variety for in each observation. Don't worry about plotting Thickness in this plot. Make sure your plot is well-labeled with appropriate axis labels and can be understood when printed in black and white.

   (b) Make an interaction plot of Year versus Locule and interpret the plot. Make sure the plot is well labeled with both a legend and axis labels.

   (c) Determine a model for the Width of the chile pods using only the 2011 data. Use a model selection technique to fit a model among the variables of Length, Thickness, Variety, Rep, and Locule. Discuss which variables should be retained and which should be dropped and whether there is evidence for any interactions between variables.

   (d) Discuss model assumptions. Perform diagnostics on your model and discuss any potential problems with the model. Plots for diagnostics can be kept in an appendix.

   (e) Write out the statistical model (in notation) with all parameters retained in the final model.

   (f) Use the model to estimate the mean width of a Casado chile pod that has length of 11 cm with 3 locales and a thickness of 2.2 mm. (You might or might not use all of these predictors depending on your final model.) Give a 95% confidence interval for this estimate. Also give a 95% prediction interval for the width of an individual chile pod with these covariates.

   (g) Test the hypothesis that there is an effect for the variety on the width of the chile pods. State the null and alternative hypotheses and do a formal hypothesis test.

   (h) Fit a new model in which the primary interest is in the shape of the chile pods, measured by the difference between Length and Width (the larger the difference, the longer and skinnier the pod is). Describe the new model and discuss any differences with the model for Width alone.

   (i) Describe in words how the number of locales (chambers) is related to the shape of the chile pods as measured by the difference between Length and Width.

2. An experiment was carried out to test the effect of two metals for pistons (metal 1 and metal 2) on the firing time of explosives. Two variables that may affect firing time are the amount of primary initiator (5, 10 and 15 mg) and packing pressure (12K, 20K, 28K psi). Each combination of metal, initiator and pressure was used in random order. The whole experiment was then repeated one more time in a different random order. See data in table 1.

Table 1: Data for problem 2

Firing time	Rep 1						Rep 2					
Initiator	5		10		15		5		10		15	
Pressure/metal	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
12000	61	57	62	67	42	73	56	56	56	64	58	73
20000	62	60	61	64	60	74	63	56	54	67	55	74
28000	66	59	60	71	59	72	57	63	59	65	58	73

Describe the statistical experimental design. Fit the simplest linear model that adequately explains the data in this experiment. List all the assumptions and explain the terms in the model. Examine the main effects, interaction terms, and do multiple comparisons of the interested treatments or combination of treatments. Make sure that you carefully assess all assumptions and write a succinct, coherent, and complete summary of your analysis.