UNM Statistics Qualifying Exam   Jan 2015   Name: _____
Due: 3 P.M., Tue Jan 13, 2015   UNM ID number

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 UNM ID number (do not include your name). 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, the more likely it is to get examined.

   Write your answers as they might appear in the methods, results, and conclusions sections of an academic paper (that is, do not include the common introduction and discussion sections). 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; consider symbols, line types, and distinct shades of gray). 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: 3 P.M., Tue Jan 13, 2015, hand-delivered to Ana Parra Lombard in the main office of the Department of Mathematics and Statistics, MSC01 1115, 1 University of New Mexico, Albuquerque, New Mexico, 87131-0001. Please do not email your solutions.

(50^{pts}) **1. Anxiolytic effect on amygdala**

To assess the clinical effects of an anxiolytic (a medication or other intervention that inhibits anxiety) on amygdala activation in females, the following preliminary experiment was conducted. Previous studies show that participants with high trait anxiety showed greater amygdala response to cued fear situations compared to those with low trait anxiety. Three female university students with mild anxiety disorder were recruited to participate. Using functional magnetic resonance imaging (fMRI) the amygdala response (unitless) under a series of anxiety trials was estimated; lower response is better. Each subject underwent three anxiolytic drug treatments (P=Placebo, L=Low, H=High), randomized in the order presented in the table below.

Data: `www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201501_pr1-DATA_Amygdala.dat`

subject	order	treatment	response
1	1	P	1.160
1	2	L	1.270
1	3	H	1.160
2	1	L	1.380
2	2	H	1.050
2	3	P	1.050
3	1	H	1.490
3	2	P	1.350
3	3	L	1.550

Analyze the data provided by this experiment. In addition to analyses and comments arising from your own curiousity, address the following as part of your write-up:

(a) What statistical design is being used, and why? Could a better design have been used, and why or why not?

(b) Is there blocking? If so, what is/are the block(s)?

(c) What is/are the nuisance factor(s) to be "averaged out" in the design?

(d) What is/are the treatment(s)?

(e) What is/are the outcome(s)/response(s)?

(f) Plot the data (not summaries of the data) in a way that helps you understand what the effects are.

(g) Write out the statistical model (in notation) and fit the model parameters.

(h) How many degrees-of-freedom are allocated to each source of variation?

(i) 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.)

(j) State and conduct statistical tests for the parameters, and interpret the test results.

(k) Are the effects of the treatment(s) and block(s) independent? Provide evidence for your claim.

(l) Report the means for the treatment groups. The preliminary data on males indicated higher activity for the low dose than for both the placebo and high dose; the same appears true for females. Write down and test the contrast comparing the low dose

response to the mean of the placebo and high dose responses. Because this test was suggested by the independent male sample, no multiple comparisons correction is required.

(m) Discuss anything else of interest. What is your conclusion about this particular anxiolytic?

(50$^{\text{pts}}$) **2. Chili pod**
The data set 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.

Data: www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201501_pr2-DATA_chili.txt

First 10 observations:

	ID	Year	Variety	Rep	FrtNo	Length	Width	Locule	Thickness
1	151	2011	jemezp	1	1	10.0	3.0	2	2.4
2	152	2011	jemezp	1	2	12.5	2.5	3	1.5
3	153	2011	jemezp	1	3	8.0	2.0	2	1.3
4	154	2011	jemezp	1	4	10.0	3.0	2	1.1
5	155	2011	jemezp	1	5	7.0	3.0	2	1.7
6	156	2011	jemezp	1	6	7.5	2.0	3	1.3
7	157	2011	jemezp	1	7	10.5	2.0	3	1.4
8	158	2011	jemezp	1	8	10.5	2.5	3	1.7
9	159	2011	jemezp	1	9	8.5	3.0	2	1.6
10	160	2011	jemezp	1	10	6.0	2.0	2	1.4

Answer the following questions.

(a) Make a plot of Width versus Length, with a separate subfigure for each number of locules 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) Is there blocking in this design? If so what are the blocks?

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

(g) Use the model to estimate the mean width of a Big Jim (bigjim) chile pod that has length of 10 cm with 4 locales and a thickness of 1.8 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.

(h) 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.

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

(j) Describe in words how the number of locales (chambers) is related to the shape of the chile pods as measured by the ratio of Length to Width.