UNM Statistics Qualifying Exam   Aug 2013   Name:_____
Due: 3 P.M., Mon Aug 19, 2013   UNM ID number

Qual Take Home (100 points) Complete both problems in this exam. It should be typed, double-spaced, no longer than three pages, no smaller than ten-point font with one-inch margins, and should be identified by your UNM ID number (do not include your name). A five-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 (in addition to summaries of data you might present with the results).

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 model equation).

4. Define and take care to assess assumptions of methods you use.

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., Mon Aug 19, 2013, 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. Vermont sheep milk production**

After several discussions, a statistian's girlfriend who worked at a small Vermont farm was convinced to conduct an experiment to assess the effect of diet on sheep milk production. Four sheep (sheep) were selected at random from the flock and marked by a numeric ear tag. Four diets (diet) were chosen to reflect typical sheep diets provided at farms in the area. Each sheep is given a different diet during each of four consecutive lactation periods (lact), and we will assume that the period between diets is sufficient so previous diets do not affect later milk production (amount).

Data: www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201308_pr1-DATA_sheepmilk.dat

	sheep	lact	diet	amount		sheep	lact	diet	amount
1	1	1	D	8.2	9	3	1	C	9.0
2	1	2	B	6.0	10	3	2	D	7.6
3	1	3	A	6.4	11	3	3	B	7.4
4	1	4	C	6.6	12	3	4	A	7.0
5	2	1	A	7.8	13	4	1	B	7.6
6	2	2	C	7.4	14	4	2	A	6.4
7	2	3	D	7.2	15	4	3	C	7.0
8	2	4	B	6.0	16	4	4	D	6.6

Analyze the data provided by this experiment and make a recommendation to the farm. In addition to analyses and comments arising from your own curiousity, please 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) Looking at the feed bags, she finds that the constituents of the diets are similar for A and B. The same is true for diets C and D. Write down and test the associated contrast.

(m) Discuss anything else of interest, and address the original goal of the experiment.

(50$^{\text{pts}}$) **2. Home prices in Eugene, Oregon during 2005**

The data are a random sample of 76 single-family homes in Eugene, Oregon during 2005 provided by Victoria Whitman, a realtor in Eugene. Realtors use experience and local knowledge to subjectively value a house based on its characteristics (size, amenities, location, etc.) and the prices of similar houses nearby. The goal of this exercise is to create a statistical model for the same purpose.

Analyze the data. Find a good predictive model for price. This should include variable selection and examination of residuals. Discuss the results of the analysis in a way that is clear to someone that does not have a lot of statistical training. In addition, what's a good asking price for a home about to go on sale of size 2.1, lot size 6, 2.1 baths, 3 beds, built in 1999, with no garages, local to Adams Elementary school?

Data: www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201308_pr2-DATA_homeprices.csv

```
Variable descriptions:
id     = ID number
price  = sale price (thousands of dollars)
size   = floor size (thousands of square feet)
lot    = lot size category (from 1 to 11)
bath   = number of bathrooms (with half-bathrooms counting as 0.1)
bed    = number of bedrooms (between 2 and 6)
year   = year built
garage = garage size (0, 1, 2, or 3 cars)
status = act (active listing), pen (pending sale), or sld (sold)
elem   = nearest elementary school (edgewood, edison, harris, adams, crest, or parker)
```

First 10 observations:

	id	price	size	lot	bath	bed	year	garage	status	elem
1	1	388.0	2.2	4	3.0	4	1940	0	sld	edison
2	2	450.0	2.1	5	3.0	4	1957	2	sld	edison
3	3	386.0	2.1	5	2.0	4	1955	2	sld	edison
4	4	350.0	1.4	6	1.0	2	1956	1	act	adams
5	5	155.5	1.8	1	2.0	4	1994	1	sld	adams
6	6	220.0	2.0	5	2.0	3	1940	1	sld	adams
7	7	239.5	1.8	4	1.1	4	1958	1	act	parker
8	8	207.0	2.3	4	2.0	4	1961	2	sld	parker
9	9	269.9	1.9	4	2.1	4	1965	2	act	parker
10	10	238.8	1.9	5	2.1	3	1968	2	sld	parker