UNM Statistics Qualifying Exam   Aug 2014   Name: ____________________
Due: 3 P.M., Tue Aug 12, 2014   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 Aug 12, 2014, 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.

(50pts) **1. Pulse rates**

Students in an introductory statistics class (MS212 taught by Professor John Eccleston and Dr Richard Wilson at The University of Queensland) participated in a simple experiment. The students took their own pulse rate. They were then asked to flip a coin. If the coin came up heads, they were to run in place for one minute. Otherwise they sat for one minute. Then everyone took their pulse again. The pulse rates and other physiological and lifestyle data are given in the data.

Five class groups between 1993 and 1998 participated in the experiment. The lecturer, Richard Wilson, was concerned that some students would choose the less strenuous option of sitting rather than running even if their coin came up heads, so in the years 1995–1998 a different method of random assignment was used. In these years, data forms were handed out to the class before the experiment. The forms were pre-assigned to either running or non-running and there were an equal number of each. In 1995 and 1998 not all of the forms were returned so the numbers running and sitting was still not entirely controlled.

Variable	Description
Height	Height (cm)
Weight	Weight (kg)
Age	Age (years)
Gender	Sex (1 = male, 2 = female)
Smokes	Regular smoker? (1 = yes, 2 = no)
Alcohol	Regular drinker? (1 = yes, 2 = no)
Exercise	Frequency of exercise (1 = high, 2 = moderate, 3 = low)
Ran	Whether student ran or sat between pulse measurements (1 = ran, 2 = sat)
Pulse1	First pulse measurement (rate per minute)
Pulse2	Second pulse measurement (rate per minute)
Year	Year of class (93 - 98)

The goal of this analysis is to understand the factors associated with pulse rate difference before and after exercise. Find a good predictive model for pulse rate difference. This should include variable selection and examination of residuals. Report the expected quality of predictions from this model, as well as any concerns you may have for prediction. Discuss the results of the analysis in a way that is clear to someone that does not have a lot of statistical training.

Consider a few relevant questions that your model should address:

1. Is there evidence that some students didn't run even though their coin toss came up heads?
2. How does Pulse1 depend on the lifestyle and physiological measurements? Are frequent exercisers fitter?
3. How does the Pulse2-Pulse1 difference depend on lifestyle and physiological measurements? Do frequent exercisers run more energetically?
4. Is there discernable differences between the years?

In addition, provide a prediction for the pulse rate difference of a student with the following characteristics (either running or not running):

Height	Weight	Age	Gender	Smokes	Alcohol	Exercise	Ran	Pulse1	Pulse2	Year
165	67	20	1	2	1	2	1			97
165	67	20	1	2	1	2	2			97

Data:

`www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201408_pr1-DATA_pulse.dat`

First few coded observations:

	Height	Weight	Age	Gender	Smokes	Alcohol	Exercise	Ran	Pulse1	Pulse2	Year
1	173	57.00	18	Female	No	Yes	Moderate	Sat	86	88	93
2	179	58.00	19	Female	No	Yes	Moderate	Ran	82	150	93
3	167	62.00	18	Female	No	Yes	High	Ran	96	176	93
4	195	84.00	18	Male	No	Yes	High	Sat	71	73	93
5	173	64.00	18	Female	No	Yes	Low	Sat	90	88	93
6	184	74.00	22	Male	No	Yes	Low	Ran	78	141	93

(50^{pts}) **2. Chinese medicinal plants**

With increased demand for Chinese medicinal plants in the U.S., a desire for locally-produced, high quality plant material is increasing, yet little is known about the feasibility of production of these species outside of China. The purpose of this study was to develop basic agronomic data for cultivation of selected Chinese medicinal plant species in northeastern United States, including the appropriate level of nitrogen for use on these crops.

Three plant species, *Agastache rugosa*, *Schizonepeta tenuifolia*, and *Leonurus japonicus*, were grown at a University Research Farm with either 0, 50, 100, or 150 kg/ha of nitrogen supplied as soy bean meal. Twenty small fields were allocated to this experiment and 18 were used. Six of the 18 small fields were randomly assigned and entirely planted with one of the three plant species Each field was then divided into quarters and each quarter was randomly assigned one of four levels of nitrogen.

The 10 center-most plants in the field quarter were harvested at the appropriate time of year. *Agastache rugosa* plants were harvested when plants were in full flower. *Schizonepeta tenuifolia* were harvested when plants were in full bloom. *Leonurus* were harvested in the fall, prior to frost. The dry weight of the 10 harvested plants were dried and their combined dry weight biomass (in grams) was measured. Of interest are differences in variety and nitrogen levels.

Data: `www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201408_pr2-DATA_plants.csv`

	species	rep	nitrogen	drywt		species	rep	nitrogen	drywt		species	rep	nitrogen	drywt
1	A.rugosa	1	0	281.8	25	L.japonicus	1	0	286.4	49	S.tenuifolia	1	0	363.6
2	A.rugosa	1	50	409.1	26	L.japonicus	1	50	318.2	50	S.tenuifolia	1	50	372.7
3	A.rugosa	1	100	454.5	27	L.japonicus	1	100	495.5	51	S.tenuifolia	1	100	427.3
4	A.rugosa	1	150	527.3	28	L.japonicus	1	150	450.0	52	S.tenuifolia	1	150	572.7
5	A.rugosa	2	0	336.4	29	L.japonicus	2	0	318.2	53	S.tenuifolia	2	0	290.9
6	A.rugosa	2	50	404.5	30	L.japonicus	2	50	404.5	54	S.tenuifolia	2	50	468.2
7	A.rugosa	2	100	368.2	31	L.japonicus	2	100	472.7	55	S.tenuifolia	2	100	600.0
8	A.rugosa	2	150	554.5	32	L.japonicus	2	150	531.8	56	S.tenuifolia	2	150	604.5
9	A.rugosa	3	0	277.3	33	L.japonicus	3	0	436.4	57	S.tenuifolia	3	0	318.2
10	A.rugosa	3	50	413.6	34	L.japonicus	3	50	563.6	58	S.tenuifolia	3	50	490.9
11	A.rugosa	3	100	440.9	35	L.japonicus	3	100	550.0	59	S.tenuifolia	3	100	572.7
12	A.rugosa	3	150	454.5	36	L.japonicus	3	150	654.5	60	S.tenuifolia	3	150	677.3
13	A.rugosa	4	0	309.1	37	L.japonicus	4	0	404.5	61	S.tenuifolia	4	0	272.7
14	A.rugosa	4	50	290.9	38	L.japonicus	4	50	586.4	62	S.tenuifolia	4	50	463.6
15	A.rugosa	4	100	509.1	39	L.japonicus	4	100	600.0	63	S.tenuifolia	4	100	404.5
16	A.rugosa	4	150	390.9	40	L.japonicus	4	150	563.6	64	S.tenuifolia	4	150	436.4
17	A.rugosa	5	0	240.9	41	L.japonicus	5	0	440.9	65	S.tenuifolia	5	0	404.5
18	A.rugosa	5	50	336.4	42	L.japonicus	5	50	450.0	66	S.tenuifolia	5	50	372.7
19	A.rugosa	5	100	536.4	43	L.japonicus	5	100	540.9	67	S.tenuifolia	5	100	390.9
20	A.rugosa	5	150	513.6	44	L.japonicus	5	150	550.0	68	S.tenuifolia	5	150	472.7
21	A.rugosa	6	0	504.5	45	L.japonicus	6	0	477.3	69	S.tenuifolia	6	0	531.8
22	A.rugosa	6	50	590.9	46	L.japonicus	6	50	636.4	70	S.tenuifolia	6	50	518.2
23	A.rugosa	6	100	713.6	47	L.japonicus	6	100	536.4	71	S.tenuifolia	6	100	731.8
24	A.rugosa	6	150	790.9	48	L.japonicus	6	150	709.1	72	S.tenuifolia	6	150	640.9

Analyze the data provided by this experiment. 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 only 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) Perform pairwise comparisons based on your final model and summarize which pairs of treatment combinations are different.

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