UNM Statistics Qualifying Exam | Jan 2013 | Name: ____________________

Due: 12 noon, Mon Jan 14, 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 introduction and discussion sections). Insert tables and figures (well-labelled and cross-referenced from text, such as, "in Table 1 ...", or if in the appendix, "in Table A1 ...") to support your points. 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.

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:** 12 noon, Mon Jan 14, 2013, hand-delivered or mailed 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. Antibiotic capsules**

This experiment involves the absorption time (minutes) of a particular type of antibiotic capsule. Four vendors (W, X, Y, Z) provide the capsule material. From each vendor's material, equal amounts are either left untreated or treated; and because of the difficulties of treating many small batches the chemical treatment in a single large batch for each vendor's material. Next, the material batches are divided into three sub-batches and formed into capsule shells by three methods (D, G, H); each method creates many capsule shells in one run. Finally, an antibiotic combination (A alone, B alone, or A and B together) is placed in each capsule shell. The shells are all stored for a period of three weeks under accelerated aging conditions, then measured. Each shell is placed in a mildly acidic solution until a proxy for absorption is indicated, at which point the time elapsed in minutes is recorded. Analyze the data, where vendors are random and the other factors are fixed, and draw conclusions.

Data: `www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201301_pr1-DATA_experiment.csv`

| | vendor | treat | formmethod | antibiotic | absorbtime | | vendor | treat | formmethod | antibiotic | absorbtime |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | W | treated | D | A | 17.2 | 37 | Y | treated | D | A | 17.8 |
| 2 | W | treated | D | AB | 17.4 | 38 | Y | treated | D | AB | 18.4 |
| 3 | W | treated | D | B | 15.8 | 39 | Y | treated | D | B | 18.4 |
| 4 | W | treated | G | A | 15.6 | 40 | Y | treated | G | A | 19.0 |
| 5 | W | treated | G | AB | 17.8 | 41 | Y | treated | G | AB | 21.4 |
| 6 | W | treated | G | B | 17.2 | 42 | Y | treated | G | B | 18.0 |
| 7 | W | treated | H | A | 17.8 | 43 | Y | treated | H | A | 18.4 |
| 8 | W | treated | H | AB | 16.6 | 44 | Y | treated | H | AB | 18.6 |
| 9 | W | treated | H | B | 17.0 | 45 | Y | treated | H | B | 16.2 |
| 10 | W | untreated | D | A | 26.2 | 46 | Y | untreated | D | A | 22.0 |
| 11 | W | untreated | D | AB | 25.2 | 47 | Y | untreated | D | AB | 23.8 |
| 12 | W | untreated | D | B | 21.8 | 48 | Y | untreated | D | B | 24.2 |
| 13 | W | untreated | G | A | 22.8 | 49 | Y | untreated | G | A | 17.4 |
| 14 | W | untreated | G | AB | 22.8 | 50 | Y | untreated | G | AB | 20.0 |
| 15 | W | untreated | G | B | 14.6 | 51 | Y | untreated | G | B | 16.0 |
| 16 | W | untreated | H | A | 27.2 | 52 | Y | untreated | H | A | 19.6 |
| 17 | W | untreated | H | AB | 26.6 | 53 | Y | untreated | H | AB | 24.4 |
| 18 | W | untreated | H | B | 23.2 | 54 | Y | untreated | H | B | 15.6 |
| 19 | X | treated | D | A | 19.6 | 55 | Z | treated | D | A | 20.8 |
| 20 | X | treated | D | AB | 21.2 | 56 | Z | treated | D | AB | 18.2 |
| 21 | X | treated | D | B | 19.6 | 57 | Z | treated | D | B | 22.4 |
| 22 | X | treated | G | A | 16.0 | 58 | Z | treated | G | A | 17.6 |
| 23 | X | treated | G | AB | 19.0 | 59 | Z | treated | G | AB | 17.6 |
| 24 | X | treated | G | B | 16.6 | 60 | Z | treated | G | B | 17.0 |
| 25 | X | treated | H | A | 16.2 | 61 | Z | treated | H | A | 15.6 |
| 26 | X | treated | H | AB | 14.8 | 62 | Z | treated | H | AB | 18.8 |
| 27 | X | treated | H | B | 14.8 | 63 | Z | treated | H | B | 16.4 |
| 28 | X | untreated | D | A | 18.8 | 64 | Z | untreated | D | A | 23.6 |
| 29 | X | untreated | D | AB | 24.6 | 65 | Z | untreated | D | AB | 22.6 |
| 30 | X | untreated | D | B | 19.8 | 66 | Z | untreated | D | B | 18.0 |
| 31 | X | untreated | G | A | 21.8 | 67 | Z | untreated | G | A | 20.8 |
| 32 | X | untreated | G | AB | 21.0 | 68 | Z | untreated | G | AB | 22.8 |
| 33 | X | untreated | G | B | 20.4 | 69 | Z | untreated | G | B | 12.0 |
| 34 | X | untreated | H | A | 26.6 | 70 | Z | untreated | H | A | 24.4 |
| 35 | X | untreated | H | AB | 26.4 | 71 | Z | untreated | H | AB | 27.2 |
| 36 | X | untreated | H | B | 27.2 | 72 | Z | untreated | H | B | 23.8 |

(50pts) **2. Diabetes**

Ten baseline variables, age, sex, body mass index, average blood pressure, and six blood serum measurements, were obtained for each of $n = 442$ diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline[1]. The first 20 observations are shown below, the variables are named: age, sex, bmi, map, tc, ldl, hdl, tch, ltg, glu, and y. Analyze the data. Find a good predictive model for disease progression (y). This should include variable selection and examination of residuals. Discuss the results of the analysis in plain English. Evaluate and address the potential for collinearity problems.

Data: `www.stat.unm.edu/~erike/exams/UNM_Stat_Exam_Qual_takehome_201301_pr2-DATA_diabetes.txt`

| | age | sex | bmi | map | tc | ldl | hdl | tch | ltg | glu | y |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 59 | 2 | 32.1 | 101.0 | 157 | 93.2 | 38.0 | 4.0 | 4.9 | 87 | 151 |
| 2 | 48 | 1 | 21.6 | 87.0 | 183 | 103.2 | 70.0 | 3.0 | 3.9 | 69 | 75 |
| 3 | 72 | 2 | 30.5 | 93.0 | 156 | 93.6 | 41.0 | 4.0 | 4.7 | 85 | 141 |
| 4 | 24 | 1 | 25.3 | 84.0 | 198 | 131.4 | 40.0 | 5.0 | 4.9 | 89 | 206 |
| 5 | 50 | 1 | 23.0 | 101.0 | 192 | 125.4 | 52.0 | 4.0 | 4.3 | 80 | 135 |
| 6 | 23 | 1 | 22.6 | 89.0 | 139 | 64.8 | 61.0 | 2.0 | 4.2 | 68 | 97 |
| 7 | 36 | 2 | 22.0 | 90.0 | 160 | 99.6 | 50.0 | 3.0 | 4.0 | 82 | 138 |
| 8 | 66 | 2 | 26.2 | 114.0 | 255 | 185.0 | 56.0 | 4.5 | 4.2 | 92 | 63 |
| 9 | 60 | 2 | 32.1 | 83.0 | 179 | 119.4 | 42.0 | 4.0 | 4.5 | 94 | 110 |
| 10 | 29 | 1 | 30.0 | 85.0 | 180 | 93.4 | 43.0 | 4.0 | 5.4 | 88 | 310 |
| 11 | 22 | 1 | 18.6 | 97.0 | 114 | 57.6 | 46.0 | 2.0 | 4.0 | 83 | 101 |
| 12 | 56 | 2 | 28.0 | 85.0 | 184 | 144.8 | 32.0 | 6.0 | 3.6 | 77 | 69 |
| 13 | 53 | 1 | 23.7 | 92.0 | 186 | 109.2 | 62.0 | 3.0 | 4.3 | 81 | 179 |
| 14 | 50 | 2 | 26.2 | 97.0 | 186 | 105.4 | 49.0 | 4.0 | 5.1 | 88 | 185 |
| 15 | 61 | 1 | 24.0 | 91.0 | 202 | 115.4 | 72.0 | 3.0 | 4.3 | 73 | 118 |
| 16 | 34 | 2 | 24.7 | 118.0 | 254 | 184.2 | 39.0 | 7.0 | 5.0 | 81 | 171 |
| 17 | 47 | 1 | 30.3 | 109.0 | 207 | 100.2 | 70.0 | 3.0 | 5.2 | 98 | 166 |
| 18 | 68 | 2 | 27.5 | 111.0 | 214 | 147.0 | 39.0 | 5.0 | 4.9 | 91 | 144 |
| 19 | 38 | 1 | 25.4 | 84.0 | 162 | 103.0 | 42.0 | 4.0 | 4.4 | 87 | 97 |
| 20 | 41 | 1 | 24.7 | 83.0 | 187 | 108.2 | 60.0 | 3.0 | 4.5 | 78 | 168 |

[1] Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani (2004) "Least Angle Regression," Annals of Statistics (with discussion), 407–499.