Final Examination

Name: 

Please indicate the sections that you attend.

Lecture:  (circle one)  Bret Larget  Jun Zhu

Discussion:  (circle one)  Sang-Hoon Cho  Xiwen Ma  Tao Yu

Instructions:

1. The exam is open book. You may use the Course Notes, other texts, lecture notes, homework solutions, your notes, and a calculator. You may not use a laptop computer.

2. Do all your work in the spaces provided. If you need additional space, use the back of the preceding page, indicating clearly that you have done so.

3. To receive full credit, you must show your work. We will award partial credit.

4. Use your time wisely. Do not dwell too long on any one question. Answer as many questions as you can in the time allowed.

5. Note that some questions have multiple parts. For some questions, these parts are independent, so you can work on part (b) or (c) separately from part (a).

For graders’ use.

<table>
<thead>
<tr>
<th>Question</th>
<th>Possible Points</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
1. Weights of ears of organic sweet corn grown are approximately normally distributed with a mean of 11.1 ounces and a standard deviation of 3.1 ounces.

(a) The lightest 8% of the ears all weigh less than what weight?

(b) What is the probability that a single ear of corn chosen at random from this population weighs more than 12.0 ounces?

(c) What is the probability that the mean weight of a random sample of a dozen ears of corn weighs more than 12.0 ounces?

(d) In a random sample of a dozen ears of corn, what is the probability that one or fewer ears of corn weigh more than 12.0 ounces?

(e) In a random sample of 288 ears of corn, use an approximation method to estimate the probability that 125 or more ears weigh more than 12.0 ounces.
2. The left ventricular ejection fraction (LVEF) measures the fraction of blood (on a scale from 0 to 1) ejected from the left ventricle per heart beat. In subjects with enlarged hearts, this fraction can be quite low. The following data represents the LVEF measurements from a random sample of \( n = 27 \) subjects with enlarged hearts and is displayed with a histogram and with a normal probability plot. If we let the \( i \)th observation be \( y_i \), two data summaries are \( \sum_i y_i = 6.05 \) and \( \sum_i (y_i - \bar{y})^2 = 0.1661 \).

0.07 0.12 0.13 0.14 0.17 0.17 0.18 0.19
0.19 0.19 0.20 0.20 0.22 0.23 0.24 0.24
0.24 0.24 0.28 0.30 0.30 0.32 0.32 0.40
0.40

(a) Find a 99% confidence interval for the mean LVEF in this population.

(b) Find a 95% confidence interval for the proportion of subjects in the population whose LVEF measurements are greater than 0.25.

(c) The plots can be used to examine model assumptions. Circle one of the summaries below and add a very brief justification of your response.

1. There is strong evidence that the population is normal.
2. There is evidence of mild skewness and non-normality in the population, but the inference procedures are robust to such deviations and the confidence intervals are still valid.
3. There is evidence of skewness and non-normality in the population so the confidence intervals should not be trusted.
3. In an environmental monitoring project near an industrial plant, sulfide concentrations are measured in six monitoring wells. There are four measurements from each well. Summary data is displayed below.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>22.88</td>
<td>22.25</td>
<td>20.55</td>
<td>24.77</td>
<td>32.95</td>
<td>41.93</td>
</tr>
<tr>
<td>variance</td>
<td>80.33</td>
<td>211.56</td>
<td>40.3</td>
<td>102.82</td>
<td>30.62</td>
<td>9.26</td>
</tr>
<tr>
<td>sample size</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

(a) Carry out a global $F$ test. You may use the fact that $SS_{\text{total}} = 2794$ to simplify the necessary calculations. Display an ANOVA table. State the hypotheses and interpret the results.

(b) Pick one assumption of ANOVA and briefly describe a means to examine the assumption with a graph of the data. What features of the graph might cause you to question the validity of the inference in (a)?
4. Continue the previous problem with the following additional information. Ground water near the industrial plant flows in a known direction. Upgradient wells A and B are located in an area that is the source of ground water near the industrial plant. In contrast, downgradient wells C, D, E, and F are located in an area where much of the ground water has passed through the location of the industrial plant. Four specific null hypotheses of interest are as follows.

1. $H_0: \mu_C - (\mu_A + \mu_B)/2 = 0$
2. $H_0: \mu_D - (\mu_A + \mu_B)/2 = 0$
3. $H_0: \mu_E - (\mu_A + \mu_B)/2 = 0$
4. $H_0: \mu_F - (\mu_A + \mu_B)/2 = 0$

(a) Use the Bonferroni method to test the hypotheses simultaneously versus one-sided alternative that the difference is positive with an experiment-wide error rate of $\alpha = 0.04$ (please note the unconventional choice).

(b) Summarize your conclusions in the context of the problem.

(c) Briefly explain how the particular tested contrasts address questions of practical and scientific importance.
Researchers are interested in studying the change in bone density as adults age. To the right is a scatter plot of $Y$, lumbar spine bone density (grams per cm$^3$), versus $X$, age (years), in a sample of 41 American women. Some useful summary statistics are below.

$$\bar{x} = 48.9, \quad \bar{y} = 0.759, \quad \sum_i (x_i - \bar{x})^2 = 5540,$$

$$\sum_i (y_i - \bar{y})^2 = 0.756, \quad \sum_i (x_i - \bar{x})(y_i - \bar{y}) = -41.9$$

(a) Find the simple linear regression line.

(b) Carry out a hypothesis test that the slope of the regression line is zero versus the alternative that it is negative. Interpret the results of this test in the context of the problem.

(c) Use the regression equation to predict the bone densities of a sixteen (16) year old girl and a fifty (50) year old woman.

(d) The full widths of the two 95% prediction intervals based on the equation on page 374 of the Course Notes are respectively 0.47 and 0.43 for the 16-year-old girl and 50-year-old woman. Briefly explain why one of these prediction intervals is much more reliable than the other.