Ground Rules:

Again. This is an “open book” “take-home” exam. You are welcome to use any reference materials you wish. You are welcome to use the computer as you wish, too. However, you MUST work this exam by yourself and you may not consult with anyone.

Instructions and Checklist

1. Start each problem on a new page.
2. Write your name on every page
3. Make a copy of your exam for safekeeping (sometimes a mailed exam is lost)
4. Submit a completed signature page.

How to submit your exam:

Dear Worcester “in-class” section,

Our last class is Monday December 7, 2015. I don’t expect that you will submit your test that day.

1. If you want to hand deliver your test, please deliver it to Linda Hollis, 4th floor Benedict Building. Please be aware that she does not work on Fridays. So you would need to get your exam to her by Thursday December 10. Please make sure that you retain a copy!! OR
2. Mail your completed exam to my home address (see below) with post-mark December 11, 2015

Online Section:

1. Upload your completed exam to the ASSIGNMENT tab no later than 11:59 pm on Friday December 11, 2015. This must be a single pdf and be named using the convention lastname_exam4.pdf. OR
2. Mail your completed exam to me with post-mark December 11, 2015 to my home address address below.

Please use my HOME address for mailing (this is because the University can be slow with mail and my turn around time for submitting grades is short). Note – if you do an overnight mailing, please do NOT require my signature. Thank you.

Carol Bigelow
59 Butter Hill Road
Pelham, MA 01002-9760
Signature

This is to confirm that in completing this exam, I worked independently and did not consult with anyone.

Signature: ____________________________________________

Printed Name: _________________________________________

Date: ____________________________
1. (10 points total)

1a. (2 points) Multiple Choice – Pick ONE (from A, B, C, D, and E)
The Student t-distribution with one degree of freedom
(1) has fatter tails than the standard normal distribution
(2) can be used to obtain confidence limits for the mean of a normal distribution from a sample of 2 observations
(3) has mean 0
(4) has variance 1

A. 1, 2 and 3  
B. 1 and 3  
C. 2 and 4  
D. Only 4  
E. 1, 2, 3 and 4

1b. (2 points) Multiple Choice – Pick ONE
An experimenter obtains, on the basis of a sample of size 10, a 95% confidence interval estimate of mean height with confidence limits equal to 66 and 74 inches. Assuming his calculations are correct, this result also depends on which of the following assumptions:

A. The mean heights are normally distributed, the sample is random and the critical values were obtained from the t-distribution.  
B. we have 95% confidence that a person’s height lies between 66 and 74 inches provided the data were randomly obtained and the critical values were obtained from a normal distribution.  
C. we have 95% confidence that the population mean height lies between 66 and 74 inches provided the data were randomly obtained and the critical values were obtained from a t-distribution.  
E. All of the above

1c. (2 points) True or False
An experimenter obtains, on the basis of a sample of size 10, a 95% confidence interval estimate of mean height with confidence limits equal to 66 and 74 inches. True or false: The sample mean is 70 inches.

1d. (4 points) Fill in the blank (show all work to receive full credit)
An experimenter obtains, on the basis of a sample of size 10, a 95% confidence interval estimate of mean height with confidence limits equal to 66 and 74 inches. The sample standard deviation is _____
2. (10 points total)

Launched in 1973, the Bogalusa Heart Study is a National Institutes of Health (NIH) prospective study of diet and cardiovascular disease risk. Among its many objectives was a comparison of Caucasians and African Americans. In one set of analyses, the average seated systolic blood pressure of the 1554 Caucasian girls was reported to 104.4 mmHg with a sample standard deviation of 9.0 mmHg; for the 927 African American girls the average seated systolic blood pressure was 104.7 mmHg with a sample standard deviation of 9.3 mmHg.

2a. (4 points)
State the appropriate formula to calculate a 95% confidence interval for the difference in mean blood pressure of the two ethnic groups from which the subjects were recruited. State any assumptions you need to use the formula.

2b. (4 points)
Using the formula you gave in question 2a, obtain the 95% confidence interval for the difference in mean blood pressure of the two ethnic groups from which the subjects were recruited.

2c. (2 points)
In 1 sentence, interpret the confidence interval you obtained in 2b.
3. (10 points total)
Suppose a survey was done this year to find out what percentage of all Americans own a bread machine. Out of their random sample of 1,000 Americans, 317 own a bread machine. The margin of error (with high confidence) for this survey was plus or minus 3%.

3a. (2 points)
What proportion of the entire American population owns a bread machine, based on these results?

3b. (8 points)
Suppose three years ago, 29% of Americans owned a bread machine. Based on the results of this current survey, what would you conclude (with high confidence) about the population of all Americans now compared with three years ago?
4. (20 points total)
The mean ± 1sd of ln [calcium intake (mg)] among 25 girls between the ages of 12 - 14 years living below the poverty level is 6.56 ± 0.64; while for 40 girls from the same age group living above poverty level is 6.80 ± 0.76

4a. (5 points)
Is there statistically significant evidence that the variances of ln [calcium intake (mg)] are unequal in the two groups (living below the poverty level versus living above the poverty level). Carry out the appropriate statistical test. In 1 sentence, interpret your result.

4b. (3 points)
What is the appropriate statistical procedure to test whether the group means are significantly different?

4c. (7 points)
Carry out your test and report your conclusion using a p-value.

4d. (5 points)
Compute an 80% confidence interval for the difference in means between the groups and interpret your result in the context of the problem.
5. (10 points total)

A study was conducted to determine if use of oral contraceptives (OC) increases the systolic blood pressure (SBP) in women between the ages of 18 and 40. The following data were obtained:

<table>
<thead>
<tr>
<th>Participant id</th>
<th>SBP when not on OC</th>
<th>SBP when on OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>115</td>
<td>128</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
<td>115</td>
</tr>
<tr>
<td>3</td>
<td>107</td>
<td>105</td>
</tr>
<tr>
<td>4</td>
<td>104</td>
<td>102</td>
</tr>
<tr>
<td>5</td>
<td>115</td>
<td>117</td>
</tr>
<tr>
<td>6</td>
<td>126</td>
<td>132</td>
</tr>
<tr>
<td>7</td>
<td>115</td>
<td>122</td>
</tr>
</tbody>
</table>

5a (4 points)
What statistical procedure would you suggest be used to construct a 90% confidence interval for the mean difference in SBP when on OC versus not. State any assumptions you need for the statistical procedure to be valid.

5b. (4 points)
Using the formula you gave in question 5a, obtain a 90% confidence interval for the mean difference in SBP when on OC versus not.

5c. (2 points)
In 1 sentence, interpret the confidence interval you obtained in 5b.
6. (10 points total) SHORT ANSWERS:
You do NOT need to perform any calculations to answer a-c.

a. (2 points)
Three researchers, Alex, Bob, and Chuck, independently select random samples from the same population. Alex has the largest sample, Bob the next largest sample, and Chuck has the smallest sample. Each researcher constructs a 95% confidence interval for π from his data. The half-widths of the three intervals are 0.015, 0.031, and 0.062. Match each half-width with its researcher.

half-width 0.015: ________ half-width 0.031: ________ half-width 0.062:__________

b. (2 points)
A researcher, Fran, selects 100 subjects at random from a population, observes 50 have over $20 in cash in their wallet. She calculates three confidence intervals for the proportion that are expected to have over $20 in cash in their wallet. The confidence levels are 90%, 95%, and 99%, and the intervals are (0.371, 0.629), (0.402, 0.598), and (0.418, 0.582). Match each interval with its confidence level.

CI: (0.371, 0.629): ______ CI: (0.402, 0.598): __________ CI: (0.418, 0.582): ________

c. (2 points)
Two researchers, Justin and Henry, work together to study what proportion of subjects have a certain genotype in the population. They each collect a simple random sample and they find that the sample proportion is 0.70. When they construct a confidence interval based on this sample proportion, Henry comes up with (0.632, 0.768) while Justin gets (0.652, 0.788). Indicate which interval has to be wrong and why.

Narrative: Disease testing
In terms of testing people for a disease, the medical community usually assumes (Null hypothesis) that you don’t have the disease until it is proven that you do have it.

d. (1 point) {Disease testing narrative} In this case, what would a type I error represent: a false positive test result or a false negative test result?

e. (1 point) {Disease testing narrative} In this case, what would a type II error represent: a false positive test result or a false negative test result?

f. (2 points) {Disease testing narrative} Which error would you rather be protected more against, a Type I or a Type II error, in this case? Explain.
7. (15 points total)

A study done by the Ohio State University Medical Center examined whether or not taking an aspirin a day would help colon cancer patients reduce the chance of getting subsequent colon polyps. 635 patients with colon cancer participated; 317 of them were randomly assigned to the aspirin group, and the other 318 patients were assigned to a placebo (non-aspirin) group. 54 patients in the aspirin group developed subsequent polyps, compared to 86 patients in the non-aspirin group. Many confounding variables were controlled for in this well-designed randomized experiment.

7a. (2 points).
State the null and alternate hypothesis for this study.

7b. (2 points).
Suggest an appropriate statistical test for the hypotheses above.

7c. (2 points).
State any assumptions you need to conduct the test above.

7d. (4 points).
Conduct your test and state your conclusion.

7e. (5 points)
In lay terms explain the type I and II errors that could occur and what would be the consequence of each.
8. (15 points total)

The following table shows data from a study on the relationship between plasma vitamin A concentration (µmol/L) and stomach cancer.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard error</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach cancer cases</td>
<td>2.65</td>
<td>0.11</td>
<td>20</td>
</tr>
<tr>
<td>Controls</td>
<td>2.88</td>
<td>No error</td>
<td>2421</td>
</tr>
</tbody>
</table>

8a. (3 points)
If we assume that the mean plasma vitamin A concentration among controls is known with no error, then what test would you suggest be used to see if there is a statistically significant difference in the mean concentration levels for stomach cancer cases and controls?

8b. (5 points)
Calculate the value of the test statistic using the data values provided.

8d. (5 points)
Obtain the p-value for your test statistic and state your conclusion in the context of this study.

8e. (2 points) – Just give this a try!
How many stomach cancer cases are needed to achieve 80% power if the true difference in mean concentration is 0.20 (µmol/L) and a two-sided test is used at a .05 level of significance.