BST 140.651 Exam 1

Notes:

- You may use your one 8.5 by 11 formula sheet.

- Please use only the basic mathematical functions on your calculator.

- Show your work on all questions. Simple "yes" or "no" answers will be graded as if blank.

- Any p-value, confidence interval, test, power or sample size calculation must be accompanied with an interpretation.

- Please be neat and write legibly. Use the back of the pages if necessary.

- Here are the point break downs: Question 1 = 30 points, Question 2 = 27 points, Question 3 = 28 points, Question 4 = 15 points. Total = 100 points.

- Good luck!

_____

signature

1. Short answer questions - 5 points each

   a. We simulate 100 data sets of 10 observations by drawing a total of 1,000 iid variates from a distribution with mean μ and variance σ^2. We take the sample mean of each of the 100 data sets and estimate the standard deviation of the resulting 100 numbers. What is that an estimate of? (explain, briefly)

   b. Will a Student's T or Z hypothesis test for a mean with the data recorded in pounds always agree with the same test conducted on the same data recorded in kilograms? (explain)

   c. What is the variance of $\bar{X} - \bar{Y}$ assuming that the means are comprised of independent groups of iid random variables? (show your work)

   d. A researcher consulting you is estimating a prevalence, p, with a 95% confidence interval of $[.15, .34]$. He claims that there is a 95% probability that the true value of p is between .15 and .34. Would you have anything to say? (explain)

   e. Another researcher consulting you is very concerned about falsely rejecting her null hypothesis. As a result, she decides to increase the sample size of her study. Would you have anything to say? (explain)

   f. What does it mean that S^2 is an unbiased estimate of σ^2. Would you expect $\log S^2$ to be an unbiased estimate of $\log \sigma^2$? (explain intuitively; do not give a formal proof; be brief)

2. Researchers are interested in a new blood test for diagnosing a viral disease. For this blood test a *positive* result is supposed to indicate the *presence* of the disease. A study found that of 50 patients who are known to have the disease, 45 are *positive* on the blood test. Of 200 individuals who are known not to have the disease, 51 are *positive* on the blood test.

   a. (9 points) Researchers are interested in whether or not the new blood test improves on the 70% specificity of the existing standard. State relevant hypotheses, perform the relevant test and calculate and interpret the associated P-value.

   b. (9 points) A person has a positive blood test result. Interpret this positive test result without knowledge of the disease prevalence. Repeat this calculation for a negative blood test.

   c. (9 points) Given that the prevalence of the disease is 10% in a subject's population with a positive test result, calculate the probability that the person has the disease.

3. Researchers studying brain volume found that in a random sample of 16 sixty five year old subjects with Alzheimer's disease, the average loss in grey matter volume as a person aged four years was .1 mm^3 with a standard deviation of .04 mm^3.

   a. (10 points) Motivate a general formula for a $100(1 - \alpha)\%$ confidence interval for this setting. Also calculate and interpret a confidence interval for the grey matter study given in the problem. Show your work.

   b. (8 points) Calculate and interpret a P-value for the hypothesis that there is no loss in grey matter volumes as people age. Show your work.

   c. (10 points) The researchers would now like to plan a similar study in 100 healthy adults to detect a four year mean loss of .01 mm^3. Motivate a general formula for power calculations in this setting and calculate the power for a test with $\alpha = .05$? Assume that the variation in grey matter loss will be similar to that estimated in the Alzheimer's study.

4. A recent Daily Planet article reported on a study of a two week weight loss program. The study reported a 95% confidence interval for weight loss from baseline of [2 lbs, 6 lbs]. (There was no control group, all subjects were on the weight loss program.) The exact sample size was not given, though it was known to be over 200.

   a. (6 points) What can be said of a $\alpha = 5\%$ hypothesis test of whether or not there was any weight change from baseline? Can you determine the result of a $\alpha = 10\%$ test without any additional calculation or information? (explain your answers)

   b. (9 points) Can you calculate a P-value for the hypothesis test from Part a using the information given? If so, report the P-value. If not, please explain what extra information you would need to do so. (show your work)