**Multiple Choice [12 pts] -- choose the ONE BEST answer for each question below by writing the corresponding letter in the blank to the left of the question.**

1. What is the measure of sampling variability?

\_\_\_\_\_\_ **A.** Mean **B.** Replicates **C.** Response **D.** Standard Deviation **E.** Standard Error

1. Which of the following is **not** a possible HA?

\_\_\_\_\_\_ **A.** x < 17 **B.** x ≠ -17 **C.**  = 17 **D.** None are possible **E.** All are possible

1. What decision should be made if the p-value < ?

\_\_\_\_\_\_ **A.** DNR Ho **B.** Reject Ho **C.** DNR HA **D.** Reject HA **E.** None of these

1. Which of the following is **not** a possible p-value?

\_\_\_\_\_\_ **A.** -0.0345 **B.** 0.9345 **C.** 1.34e-6 **D.** None are possible **E.** All are possible

1. What type of error occurs when a true Ho is rejected?

\_\_\_\_\_\_ **A.** Type I **B.** Type II **C.** Type III **D.** Rejection **E.** No error

1. Which symbol or word represents the probability of rejecting a false Ho?

\_\_\_\_\_\_ **A.**  **B.**  **C.**  **D.** Inference **E.** Power

1. What happens to the probability of a Type II error if  is increased?

\_\_\_\_\_\_ **A.** Decreases **B.** Stays the same **C.** Increases **D.** Becomes zero **E.** Goes to ∞

1. What level of confidence should be used if =0.1 and HA: >o?

\_\_\_\_\_\_ **A.** 0.10 **B.** 0.50 **C.** 0.90 **D.** 0.95 **E.** 0.99

1. What type of confidence region is constructed if =0.1 and HA: >o?

\_\_\_\_\_\_ **A.** Interval **B.** Point estimate **C.** Upper bound **D.** Lower bound **E.** Rejection region

1. What happens to the margin-of-error if the sample size is increased?

\_\_\_\_\_\_ **A.** Decreases **B.** Stays the same **C.** Increases **D.** Becomes zero **E.** Goes to ∞

1. A t-distribution is \_\_\_\_\_\_\_\_\_\_\_ than a Z-distribution?

\_\_\_\_\_\_ **A.** narrower than **B.** same width as **C.** wider than **D.** twice as wide as **E.** taller than

1. Which t\* corresponds to a 95% confidence interval and 100 df?

\_\_\_\_\_\_ **A.** -2.365 **B.** -1.984 **C.** +1.660 **D.** +1.984 **E.** +1.984

**Short (Paragraph) Answers -- Answer THREE of the following questions with complete sentences on a separate sheet of paper. Circle the questions below that you have chosen to answer and make sure to clearly label your answers on the separate sheet. Each question is worth 3 points.**

1. What are the two major methods or techniques for performing statistical inference? Completely describe the differences between the objectives for the two methods.
2. Describe choices that you, as a researcher, can make to reduce . Which is the best choice to make and why?
3. Describe choices that you, as a researcher, can make to reduce the margin-of-error. Which is the best choice to make and why?
4. What is the difference between “statistical significance” and “practical significance”? Explain why it is important to make a distinction between these two forms of significance?
5. Completely describe the similarities and differences between a Z- and a t- distribution.

**Answer the following questions with properly labeled figures and tables, as appropriate. Make sure to start with a blank RStudio (except for your script for loading the data in 20) and MS-Word document. You can type your answers in Word or print the results from R and hand-write your answers. Save your documents regularly.**

1. **[10 pts]** Use the following information to answer the questions below: HA: >190, =0.01, =20, n=40, andx=200.
2. Define the p-value specifically for this situation, compute the p-value, and use the p-value to make a decision.
3. Compute and fully interpret the confidence region.
4. **[3 pts]** Biologists often visually estimate the number of geese in a flock to determine relative abundance. Some researchers believe, however, that biologists tend to over-estimate the true size of the flock with their visual estimates. To test this hypothesis, researchers will show a slide with a certain number of geese on it to a random sample of biologists and ask each biologist to estimate the number of geese on the slide. Assume that the standard deviation for the estimates is 40, the researchers want the mean to be estimated within 5 birds, and that they will use a 10% significance level. What sample size of biologists should the researchers use?
5. **[15/17 pts]** Last year I planted 400 ever-bearing strawberry plants. The company from which I bought the plants claimed that each plant would produce an average of 12 berries in the year following being planted. I was surprised by this claim and hypothesized that the plants would actually produce less, on average, than what the company said. To test this claim, I randomly selected 50 plants on which I counted the number of ripe berries produced for the entire season following the year that the plants were planted. Use the data in Strawberries.txt to perform an appropriate hypothesis test, at the 10% level and assuming that the standard deviation is 5 berries.

# **11 Steps for any Significance Test**

**1.** **[1]** state the rejection criterion (),

**2.** **[2]** state the null and alternative hypotheses to be tested and define the parameter(s),

**3.** **[1]** determine which hypothesis test to use – thoroughly explain why,

**4. [1]** collect the data (address type of study and randomization),

**5.**  **[2]** check all necessary assumptions (explain how you tested the validity),

**6.**  **[1]** calculate the appropriate statistic(s),

**7. [2]** calculate the appropriate test statistic,

**8.**  **[2]** calculate the p‑value,

1. **[1]** state the rejection decision,

**10**. **[2]** summarize your findings in terms of the problem, and

**11. [2]** if rejected H0, compute a 100(1-)% confidence region for the parameter.

**library(NCStats)**

**distrib(x,mean=##,sd=##,lower.tail=XXXXX,type=”X”)**

where **x** is replaced with the value of the quantitative variable or the area

**mean=##** has ## replaced by the value of the mean

**sd=##** has ## replaced by the value of the standard deviation or standard error

**lower.tail=XXXXX** has XXXXX replaced with TRUE (default) for a “left-of” and FALSE for a “right-of” calculation

**type=”X”** has X replaced with p (default) for a forward and q for a reverse question

**sqrt(x)**

**( z1 <- z.test(df$var,sd=##,mu=##,alt="XXX",conf.level=0.##) )**

where **df$var** has df replaced by the name of the data frame and var replaced by the name of the quantitative variable

**mu=##** has ## replaced by the null hypothesized mean

**sd=##** has ## replaced by the value of 

**alt="XXX"** has **"**XXX**"** replaced by **"**less**"**, **"**greater**"**, or **"**two.sided**"** (default)

**conf.level=0.##** has ## replaced by the level of confidence