Math 116 **ACTIVITY 9:** Hypothesis testing on the mean: setup, calculation, conclusion

Why

Testing to see whether a sample gives evidence of a difference or change in a parameter is one of the [two] principal methods of inferential statistics. Setting up the test from a question and data set is the critical first step, and is closely tied to calculation of the test statistic. You need to become familiar with the steps for setting up and carrying out a test so that we can move on to using the process to make decisions.

LEARNING OBJECTIVES

1. Be able to carry out a significance test for the mean of a population, and state and interpret the correct conclusion.
2. Be able to use the table of values for Student's t to determine the (range for) the P-value in a significance test on a mean.
3. Understand the risk involved in each type of decision.
4. Work as a team, using the team roles.

CITERIA

1. Success in working as a team and in fulfilling the team roles.
2. Success in involving all members of the team in the conversation.
3. Success in completing the exercises

RESOURCES

1. The Math 116 Statistics Handbook - chapter 7 on" Hypothesis Testing on the Mean" and the table of "Critical Values for Student's t".
2. The handout Hypothesis testing on a population mean from class Tuesday 4/16
3. Your calculators
4. The team role desk markers (handed out in class for use during the semester)
5. 40 minutes

PLAN

1. Select roles, if you have not already done so, and decide how you will carry out steps 2 and 3 (5 minutes)
2. Work through the group exercises given here - be sure everyone understands all results & procedures(25 minutes)
3. Assess the team's work and roles performances and prepare the Reflector's and Recorder's reports including team grade (5 minutes).

DISCUSSION
The outline of a test on a population mean:

1. Identify the variable and population and state the null hypothesis H_0 and alternative hypothesis H_A
   (Test is designed to see if we can reject H_0 and support H_A)
   For Mean:
   H_0 is always of the form $\mu =$ particular number (usually called μ_0)
   H_A can be any of the three forms $\begin{cases} \mu > \mu_0(\text{“} >'') \\ \mu < \mu_0(\text{“} <''), \\ \mu \neq \mu_0(\text{“} \neq'') \end{cases}$
2. Identify and calculate the value of the test statistic (from the data and the value used in H_0)
   for Mean: sample $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with degrees of freedom $df = n - 1$
3. Find the P-value - the probability, treating H_0 as true, of the test statistic weighing as heavily against H_0 (if only chance variation was at work) as it does for the data we have [See "notes on using the t-table"]

4. State a conclusion. If the P-value is small enough (rule of thumb: $P < .05$"significance level .05") "We have evidence of a difference (be specific)" If not "We do not have evidence of a difference (be specific)"

Notes on using the t-table to get a P-value:

We are always interested in "how far from 0 is our sample t"—and sometimes in the *direction* of the difference (larger or smaller). The exact action depends on the form of the alternative hypothesis.

1. For a ">" alternative: only *positive* values of the sample t will be significant (give evidence that the real mean is greater than μ_0). We read across the row for the correct degrees of freedom until we find numbers that bracket our sample t . The P-value is between the values at the tops of the two columns. [If sample t is larger than all values in the row, then $P < .0005$ and we have strong evidence of a difference. If sample t is smaller than all values in the row — especially if it's negative — then $P > .25$ and we dont have evidence for anything.]
2. For a "<" alternative: Only *negative* t-values will be significant (give evidence that the real mean is less than μ_0). A positive value for sample t will give $P > .25$ — no evidence for anything. We read the table as if all the t-values (numbers in the body of the table) are negative, and [as in the first situation] read across the row until we find numbers that bracket our sample t . The P-value is between the values at the tops of the two columns. [If we have to go off the right-hand edge of the table — values further from 0 —then $P < .0005$ and we have strong evidence of a difference. If we are off the left side, especially if sample t is positive, $P > .25$ and we don't have evidence of anything.
3. For a "$\neq$" alternative: *Either positive or negative* values of the sample t can be significant (give evidence that the real mean is different from μ_0). We *ignore the sign* on the sample t, and (as before) read across the appropriate row (for correct degrees of freedom). This time, though, because there are two possibilities being combined, we must double the P-value at the top of the column (in effect, adding together probabilities for "smaller" and for "larger").

Error and risk: "Statistics means never having to say you're certain"

Whatever data we have, and whatever decision we make, the result is never certain. There are two types of error that can occur (though in any specific case, only one is possible):

I When we say "Yes, we have evidence of a difference" there probably is a difference — but we may be misled by an unusual sample [this will happen a certain amount of the time] and be making a **Type I error**. The risk of a Type I error is often called "alpha risk". The "significance level" of a test (the cutoff we use for "small" P-values) is the maximum level of alpha risk we consider acceptable.

II When we say "No we do not have evidence of a difference" it is likely that the difference is small (there may even be no difference at all) but we may be misled by an unusual sample [this will happen a certain amount of the time] and be making a **Type II error**. The risk of a type II error is often called "beta risk"
Whichever conclusion we draw, there is a risk of error. "Statistics means never having to say you are certain.'

MODELS :

1. A biology student wishes to decide whether the mean length of earthworms on the Saint Mary's campus is less than 12.4 cm. She has reason to believe that the lengths are approximately normally distributed. She obtains a sample of 20 earthworms, giving a mean length 11.7 cm, with standard deviation 1.4 cm.

    1. The variable is $X =$ "length (cm) of a Saint Marys earthworm"
    $H_0 : \mu = 12.4$ [mean length for Saint Marys earthworms is 12.4 cm]
    $H_A : \mu < 12.4$ [mean length for Saint Marys earthworms is less than 12.4 cm]
    2. sample information: $n = 20, \bar{x} = 11.7, s = 1.4$
    so sample $t = \dfrac{11.7 - 12.4}{1.4/\sqrt{20}} = 2.236(df = 19)$
    3. Using the t-table [Case 2 in "Notes on using the t-table" —the "<" alternative—with $df = 19$] we find $-2.2539 <$ sample t< 2.093 , so $.01 < P < .025$
    4. Yes, the data give good evidence ($.01 < P < .025$) that the mean length of SMC earthworms is less than 12.4 cm.

[What the test says is, that if the mean really was 12.4 cm, a sample like this one would show up in less than 2.5% of the cases— so this sample is mostly inconsistent with a population mean length 12.4 cm (or more)]. We are exposed to a risk of Type I error but the risk is small.

2. Is the mean weight for left-handed boa constrictors different from the mean weight all boa constrictors? The mean weight for all boa constrictors is 12.7 Kg; a sample of 15 left-handed boa constrictors gives a mean 12.0 Kg and standard deviation 1.4 Kg. (Assume weights are approximately normally distributed).

   1. The variable is $X =$ "weight (Kg) of a left-handed boa constrictor"
      $H_0 : \mu = 12.7$
      $H_A : \mu \neq 12.7$ [We just want to know if the mean is *different*—not asking "larger", not asking "smaller"]

   2 Sample $t = \dfrac{12.0 - 12.7}{1.4/\sqrt{15}} = -1.936, df = 14$

   3. Using the t-table [Case 3 in "Notes on using the t-table" —the "$\neq$ " alternative– with $df = 14$] we ignore the sign on sample t and find $1.761 < |\text{sample } t| < 2.145$; since we must double the p-values from the top of the columns, we find $.05 < P < .10$.

   4. No, the data do not give evidence at the 5% level ($.05 < P < .10$) that the mean weight of left-handed boa constrictors is different from 12.7 Kg.

   [What the test says is that if the mean weight really is 12.7 Kg, a sample like this would show up, just by chance, between 5% and 10% of the time— this sample is reasonably consistent with a population mean weight of 12.7 Kg] We are exposed to a risk of Type II error—but the risk of Type I error would be unacceptably large if we said "yes" with this sample.

EXERCISES

1. A researcher studying the effects of nutrition on development keeps a sample of 27 young rats on a diet with reduced calcium; they show a mean weight gain 236 g , with standard deviation 52 g. For rats of this type on a normal diet, the mean weight gain is 255 g (weight gains are approximately normally distributed). Do these data show that the mean weight gain for rats on a reduced calcium diet is lower than for those on a normal diet? [Will use Case 2 — "<"— in "Notes on using the t-table"

2. A random sample of 15 households in a small city gives the following number of hours of television watched in one week (for each) . A marketing firm wishes to determine whether the mean amount of television watching per week is more than 8 hours. Does the study give evidence that the mean number of hours watched is more than 8? (Assume that number of hours watched is approximately normally distributed) [Will use Case 1 —">"— in "Notes on using the t-table"]
   # hours of television watched (n = 15)
   7.5 6.3 7.2 5.0 7.9 10.7 8.9 8.8 10.3 8.8 9.5 9.5 6.1 9.4 8.4

3. A pharmaceutical manufacturer forms tablets (pills) by compressing a granular material that contains the active ingredient and various fillers. The hardness of a sample from each lot (batch) of tablets is measured in order to control the process. The hardness should average 11.5 units. The 20 hardness values below are obtained from a random sample of 20 lots. Knowing that the hardness values are roughly normally distributed, do these values indicate that the mean hardness is different from 11.5 units (so the process needs adjustment)? [Will use Case 3 — "$\neq$" in "Notes on using the t-table"]
   hardness for 20 lots :
   11.64 11.63 11.50 11.63 11.38 11.40 11.61 11.47 11.57 11.49 11.40 11.75 11.50
   11.54 11.45 11.51 11.59 11.65 11.49 11.56

READING ASSIGNMENT (in preparation for next class)
Read chapter 8 - inference on a population proportion

SKILL EXERCISES:(hand in - individually - with assignment for this week)
Exercises for Chapter 7#4 - 9