Activity #21: Hypothesis Testing (t-test)

1: State two competing hypotheses *about the population* (we will initially assume the null hypothesis is true).
  a. H_0: null hypothesis ("dull hypothesis" – typically a statement of no effect)
  b. H_a: alternative hypothesis (the one you actually hope to demonstrate)

2: Identify potential decision errors; decide upon the level of significance (α)

3: Sketch the sampling distribution. Calculate the observed test statistic and compare it to the critical test statistic (from a table). Calculate the p-value.

4: Decide whether the *p-value* is small enough to convince you that the sample results didn't happen by chance alone

5: Make a decision and explain what your decision means
  a. Reject H_0 (the treatment had an effect)
  b. Fail to reject H_0 (we cannot conclude that the treatment had an effect)

6: Calculate the power of your study

Situation: In the 1980s, many companies experimented with "flextime," allowing employees to choose their schedules within broad limits set by management. Among other things, flextime was supposed to reduce absenteeism. Suppose a firm is willing to try flextime for a year to see if absenteeism is significantly increased or decreased. Based on data for the past decade, this firm knows its employees have averaged 6.3 days off from work. During the trial flextime year, management randomly sampled 100 employees and found they averaged 5.5 days absent (with a standard deviation of 2.9 days).

1) State the null and alternate hypotheses. Is this a one-tailed (confirmatory) or two-tailed (exploratory) test?

2) Express the consequences of both Type I and Type II errors. Which type of error has the most serious consequences? At what level should α be set?

3) All we know about the distribution of “days absent” is that $\mu = 6.3$ and $\sigma = 2.9$. We don’t know anything about the shape of the distribution. Given that we sampled 100 employees, can we sketch the sampling distribution of the sample means? What kind of distribution is this? Sketch the distribution and specify its parameters.

4) Suppose we set $\alpha = 0.05$. This means we need to shade in 5% of the distribution. Find the critical value(s) for this distribution and shade them in.

5) Locate our observed sample mean on the distribution. Convert this to the t-scale.

6) Calculate the p-value from this study. What can we conclude from this study?

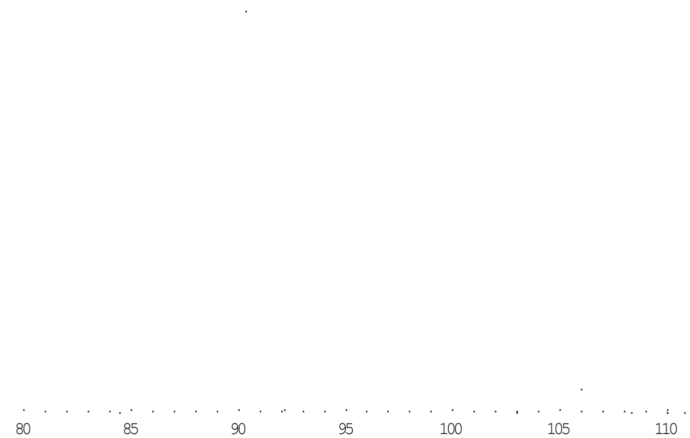
7) Sketch the sampling distribution under the null hypothesis for this study. Draw an arrow to identify the observed sample mean. Convert this mean to the t-scale. Find the critical values for $\alpha = 0.001$, 0.01, 0.05, and 0.10. Shade in these critical regions. What happens as we increase alpha? Does our decision change?

Rules: If p-value $\leq \alpha$, we can reject the null hypothesis. The result is "statistically significant"
If p-value $> \alpha$, we fail to reject the null hypothesis. The result is not "statistically significant"

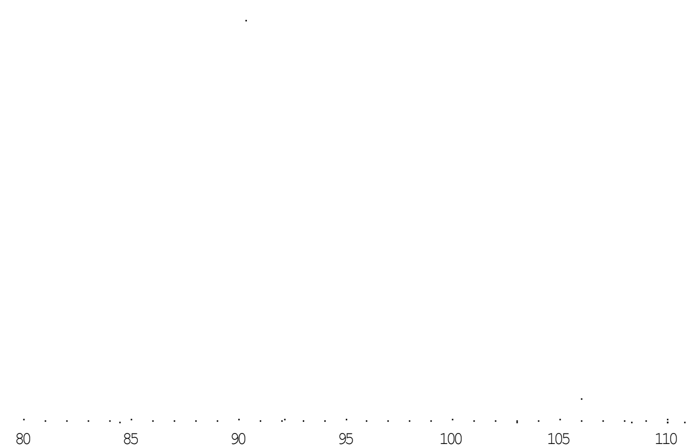
8) Suppose we find out (through divine intervention) that flextime reduces absenteeism by 1.3 days (in other words, the true state of nature is that our flextime absenteeism is actually 5 days). Given this new reality (notice the null hypothesis is no longer true), compute the power of our study.

I'll give you more practice with hypothesis tests next class. For now, let's examine the concept of power a little more in-depth. Specifically, let's find ways to increase power.

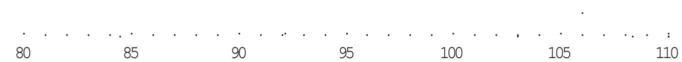
A null and a (true) alternate hypothesis are displayed below. Using an α = 0.05, the critical value is also displayed. Shade in alpha and beta. Locate the graphical representation of power.



1. What happens if we decrease alpha to α = 0.001?



2. What happens if we increase alpha to α = 0.10?



3. What happens if we increase our sample size? First, we know the standard error would decrease and our distributions would become skinnier.