

Lab 17

Hypothesis Testing and the Ashe Center Student Wellness Survey

Before starting this lab, you should be familiar with: hypothesis testing for a mean, know the terms significance level, p-value, one-sided hypothesis, two-side hypothesis and be able to state the null and alternative hypotheses for a statistical test.

About the Data

The dataset that we will be using in this lab comes from the Ashe Center Student Wellness Survey.

Load this dataset into *Stata*

```
. use http://www.stat.ucla.edu/labs/datasets/ashetest.dta
```

There should be five variables in the dataset:

weight:

Survey respondents self-reported weight. Weight is a numerical variable.

smoker:

Smokers in the survey are represented with a “1”, while non-smokers are represented with a “0”. This is a categorical variable.

howmany:

The survey asks “During the academic year, about how many times would you say you have gotten drunk after drinking?” The responses are recorded as numerical values.

gender:

A “0” represents female, while a “1” represents a male.

numpart:

The survey asks, “How many sexual partners have you had?” The responses are recorded as numerical values.

One Sample T-test

Using a t-test, we want to find out if UCLA students, on average, weigh the same as the national average. Suppose that the average weight of young adults in the United States, ages 18-22, is 160 pounds. Type:

```
. summarize weight
```

Question 1: What is the mean weight of students at UCLA? Is it different from the national average?

You’ll notice that in fact the average weight is not 160. But is it close? That is, is the sample average different from 160 because of random chance, or because UCLA students weigh less than the national average? The t-test to the rescue!

Question 2: State the null and alternative hypothesis.

Question 3: State the assumptions behind the t-test.

```
. ttest weight = 160
```

The results give first the summary statistics for the variable weight, the null hypothesis, weight equals 160, (Ho: mean(weight) = 160), followed by all three possible alternative hypothesis; weight is less than 160, (Ha: mean < 160), weight is not equal to 160, (Ha: mean \neq 160), weight is greater than 160 (Ha: mean > 160), and their respective t-statistics and p-values.

Question 4: Interpret the t-statistic and p-value for the null and alternative hypothesis stated above.

Question 5: What do you conclude about the weights of UCLA students compared with the national average of 160 lbs?

Two Sample T-test

Is there a difference in the weight of those who smoke and those who do not smoke? This calls for a test to see if the difference between two groups is significant. We will again use a t-test, but this time it will be a two-sample t-test, this means that we will test if the difference between the two groups is less than, equal to, or greater than zero. First let's look at the summary statistics for weight classified by smoker.

```
. sort smoker  
. by smoker: summarize weight
```

Question 6: Does it appear that smokers weigh the same as non-smokers from comparing their means?

Question 7: State the null and alternative hypothesis.

```
. ttest weight, by(smoker) unequal
```

The “unequal” means that this t-test does not assume that the standard deviations of the two populations (smokers and non-smokers) are equal. The output is much the same as for a one-sided t-test. The summary statistics for both groups are given, as well as the difference (diff) between smokers and non-smokers, followed by the null hypothesis (Ho: mean(0) - mean(1) = diff = 0) and the three possible alternative hypothesis.

Question 8: Interpret the t-statistic and p-value for the null and alternative hypothesis stated above.

Question 9: What do you conclude about the difference in weight between smokers and non-smokers?

Here is a list of *Stata* commands we used in this hypothesis testing lab. Use the space next to each command to make notes on what that command does.

```
summarize
```

```
ttest
```

```
ttest, unequal
```

```
sort
```

```
ttest (var1), by(var2)
```

Assignment

Using the other variables in the data set answer the following two questions. Please give the null and alternative hypothesis, type of test needed, test statistic, p-value and a well-stated conclusion:

Question 10: Do men drink more than women?

Question 11: Is there are difference in promiscuity between men and women?