As a warm-up for this lab session, you should look at the tutorials for beginners, one mean, two means, and paired means.

In this lab session you will be working with county-level data on U.S. private land conservation. The counties are a randomly selected subset (20%) of all counties in the continental U.S. The data are contained in the 'main_data' datasheet of the landcons_sample.xls file. The table gives variable names and definitions.

Variable	Definition
fips	A unique numerical code identifying state and county
state	The 2-letter state abbreviation
county	The county name
pprivcons00	The percent of private land in a county owned or held in conservation easements by land trusts in 2000
pci00	The per-capita income of a county's residents in 2000
hpci00	= 1 for the counties in the upper quartile of the distribution of pci00 (i.e. the richest 25% of counties in the sample)
amscale	A composite score for the quality of natural amenities in a county. The score increases with increases in a county's water abundance, topographical variation, and average winter temperature. The score decreases with increases in summer average temperatures and humidity.
hamscale	=1 for the counties in the upper quartile of the distribution of amscale (i.e. the nicest 25% of counties in the sample)
ppub00	The percent of county land owned by the federal government (e.g., USPS, USFS, etc.)
hppub00	=1 for the counties in the upper quartile of the distribution of ppub00 (i.e. the 25% of counties with the most public land in the sample)

Once we upload these data into JMP, we will work through 1-6.

1. Your friend bet you in 1998 that the mean per-capita income across U.S. counties would not exceed $22,000 by the year 2000. Unfortunately, you only have access to the random sample of U.S. counties provided here. Because your bet cannot be resolved with perfect certainty, you have agreed to resolve it with the method of hypothesis testing using this sample. You have generously agreed to let your friend's claim be considered the null hypothesis, and you have agreed to use a significance level of 0.05.

    a. What is H_O, the null hypothesis?

b. What is H_A, the null hypothesis?

c. Is this a one or two-tailed test?

d. What is the mean amount of per-capita income in 2000 from our sample?

e. Does the sample mean differ significantly from $22,000? Who wins the bet?

2. Your friend also believed that the average level of private land conservation in a U.S. county would not exceed 0.30 % of the private land in a county in 2000. You disagreed on this and bet your friend that she was wrong. You have agreed to let your position stand as the null hypothesis.

   a. What is H_O, the null hypothesis?

   b. What is H_A, the alternative hypothesis?

   c. What is the mean value of `pprivcons00`?

   d. Does `pprivcons00` look normally distributed? Contrast this with the distribution of `pci00` – does this variable look normally distributed? Given the distribution of `pprivcons00` can we use the same procedure as in #1?

   e. Who wins the bet?

3. Another friend claims that private land conservation is a luxury that only rich counties can afford. You think land conservation decisions are complicated, and have never thought there is a systematic relationship with income. You decide to initially evaluate your prior beliefs using hypothesis testing. The two variables you have at your disposal are `hpci00` and `pprivcons00.`

   a. What is H_O, the null hypothesis?

   b. What is H_A, the null hypothesis?

c. Is this a one-sample or two-sample t-test? Two-tailed or one-tailed?

d. What do you conclude?

4. An alternative view is that private conservation is simply greater in counties that are endowed with an abundance of natural amenities. Test whether this is the case using the same procedure as in #3, but now the two variables at your disposal are `hamscale` and `pprivcons00`. What do you conclude?

5. Finally, another view is that private conservation is a substitute for public land conservation. Test whether this is the case using the same procedure as in #3, but now the two variables at your disposal are `hppub00` and `pprivcons00`. What do you conclude? Does this invalidate your conclusion?

6. Using the tools of two-sample comparisons of means and t-tests can you ascertain whether natural amenities or incomes or public conservation or other factors are causing higher levels of land conservation in different counties?

We may be able to better isolate the relationship of per-capita income and private land conservation by using two-sample *t* procedures with paired data. To create the paired data, we randomly draw an observation from each state from the sample where `hpci00`=0. We also randomly draw an observation from each state from the sample where `hpci00`=1.

Doing this generates the datasheet labeled 'paired_data' in the landcons_sample.xls file. Notice that we have paired data from the 36 states with at least one low-income and high-income county. The following table defines the variables in the 'paired_data' sheet.

Variable	Definition
statepair	The 2-letter state abbreviation that contains the randomly drawn pair of counties
hpci_pprivcons	The value of pprivcons for a randomly drawn county from the sample of high-income counties
lpci_pprivcons	The value of pprivcons for a randomly drawn county from the sample of low-income counties

Once we upload these data into JMP, we will work through #7.

7. Using the paired data, is there evidence that the mean amount of private land conservation is greater in high income counties? What advantages does this test have over the unpaired comparisons of means in #3?