Comparing Two Groups

Part I: Before Class

Objective: This lab will give you practice examining differences between two groups of data. We will also analyze the significance of the test results in the context of the studies presented in this assignment.

Before starting this lab, you should…

1. Read this assignment all the way through;

2. Know:

   - The concept of the null and the alternative hypotheses;

   - How to apply t-tests;

   - When to use 1-tailed or 2-tailed tests;

   - How to interpret p-values.

Part II: The Data for the Take Home Section

Risk Perception

Psychologists are interested in finding ways of measuring perception of risk since it is an important component in any decision-making process. The data provided here come from

611 participants who were asked to provide a judgment of risk on several activities using a scale of 0-100 (100 being high risk). In this lab assignment, we compare how two different groups, namely, men and women, perceive risks based on five different activities.

Source: Lisa K Carlstrom, J. Arthur Woodward, and Christina G.S. Palmer. "Evaluating the Simplified Conjoint Expected Risk Model: Comparing the Use of Objective and Subjective Information," *Risk Analysis* 3(2000) v. 20: 385-92.

Comparing Two Groups — V.1

- You must open the web page below and double click on "Risk Stata." Download the file called "risksmall.dta".

http://www.stat.ucla.edu/~rgould/m12s01/risksmall.dta

Once you have downloaded it, you should be able to see the seven variables in that dataset.

SUBID

- ID of subject (or participant)

GENDER

- 0 = Female, 1 = Male

APPL

- How risky it is to use household appliances, on a scale of 0-100 (0 = low, 100 = high)

NUC

- How risky it is to live near a nuclear power station, on a scale of 0-100 (0 = low, 100 = high)

POOL

- How risky it is to swim in indoor public pool each weekend, on a scale of 0-100

(0 = low, 100 = high)

PLANE

- How risky it is to fly on a commercial airplane every month, on a scale of 0-100

(0 = low, 100 = high)

XRAY

- How risky it is to receive diagnostic x-rays every six months, on a scale of 0-100 (0 = low, 100 = high)

- For more information on this case study, please refer to the explanation of the data posted on the following web address:

http://www.stat.ucla.edu/projects/datasets/risk_perception.html

Comparing Two Groups — V.1

Take-home questions:

- Answer the first two questions before you examine the data.

1. Which activity, if any, do you think will show differences in gender? Which activities do you think men will tend to find riskier than women? Which one(s) will both find to be equally risky?

2. How risky would <u>you</u> rate each of the five activities, on a scale of 0-100 (0 = low, 100 = high)?

3. Use graphical techniques (whichever ones you find helpful) to compare risk ratings from men and women in each of the activities. Interpret the graphs. Any surprises?

4. A critic might say that the differences in risk perception between men and women are due to chance. Do a statistical test for each of the five activities. Which significance level will you use? Will you assume the standard deviations to be the same for both populations? Explain your options in detail.

5. How did your predictions of gender differences in question "1" compare to the results in question "4"?

6. Compare and discuss all the results you have obtained for men and women.

Part III: Lab Exercises

These are the Stata commands you may find useful to do this assignment.

insheet using *filename* **-** reads unformatted ASCII (text) data. It is intended for reading files created by spreadsheet. This command is especially useful when you don't know the variables (their amount and names) of a dataset.

sort x **-** arranges the observations of a variable in ascending order of values. There is no limit to the number of the variables in the data. Missing values are interpreted as being larger than any other number and thus are placed last.

> **display tprob**(df**,**t**)** - you need to specify the degrees of freedom and the
>
> t-statistics. This function returns the p-value of the sample(s) test by performing a 2-tailed
>
> t-distribution.
>
> **display tprob**(df,t)**/2** - returns the p-value of the sample(s) for an upper-tail
>
> t-distribution.
>
> **display 1 - tprob**(df**,**t)**/2**- returns the p-value of the sample(s) for a lower-tail
>
> t-distribution.
>
> **ttest -** performs one- and two-sample test on the equality of means. There are several
>
> variations of this command, the most common ones are:

ttest x = # **-** tests the hypothesis that x has a mean of #.

> **ttest** x**, by**(z) - performs a test of x controlling for z.
>
> **ttest** x**, by**(z) **unequal** - same above. Except that the word "unequal" indicates that the
>
> two-sample data are not to be assumed to have equal variances.

use *filename* **-** loads a Stata-format file dataset from the Web. If filename is specified without the Stata extension, ".dta" is assumed.

Activity 1: Hypothesis testing

We will be using the dataset collected by Prof. Gould in his Statistics 12 class, in Spring 2000. First, you should download the dataset from the following address:

http://www.stat.ucla.edu/~rgould/datasets/m12s00.dta

Comparing Two Groups — V.1

7. Suppose that the average weight of the young adults in the 18-22-age range in the United States is 150 pounds. We want to know whether the sample we have (Prof. Gould's students) is from that population. In order to test this hypothesis we need to perform a one-sample t-test. Thus, type:

ttest weight = 150

Questions:

a. State the null and the alternative hypothesis (two-sided): (1) in words and (2) using mathematical notation.

b. Interpret the t-value and the confidence interval. What do you conclude about the null hypothesis?

c. Using the mean and the standard deviation of the sample, hand calculate the standard error, and the 95% confidence interval. Compare the values you found with the values given on the printout.

2) Next, we want to examine whether women have lower levels of math anxiety than men. If we assume the data represent independent random samples of men and women, we can apply a two-sample t-test. Also, notice that we need to perform a one-tailed t-test. First, we have to sort the variables by gender, as follows:

sort gender

Now, run the t-test.

ttest mathanxi, by(gender)

Questions:

a. State the null and the alternative hypotheses using your own words and mathematical notation.

b. Using the mean and the standard deviation of the sample, hand calculate the standard error, the t-value, and the 95% confidence interval. Compare the values you found with the values given on the printout.

Comparing Two Groups — V.1

Now, let's run another two-sample t-test. This time, let's assume that the two populations' variances are unequal. Thus, type:

ttest mathanx, by (gender) unequal

Note that the degrees of freedom are different from the ones in the previous test

c. Consider the standard deviation in each of the samples we are examining. What do you think would be more reasonable to assume; that these two populations' variances are *equal* or *unequal*? Explain.

d. Back to our research question. What is your final conclusion about the question of whether women tend to have lower levels of math anxiety than men? Explain your answer in detail and take into consideration the results you obtained assuming *equal* variances and those assuming *unequal* variances.

e. Do you think the sample is representative of the population of all college-age men and women in the U.S.?

Activity2: Choosing "degrees of freedom" for t-tests in Stata

Stata uses a specific method (the Satterthwaite's equation) to compute the degrees of freedom. This might not be the same method you learned in this class. Notice that the way you choose to calculate the degrees of freedom of the test affects its p-values. The good news is that you can use Stata to find the p-values for a t-statistic with any degrees

of freedom. The command you will need to use will depend on the hypothesis you are testing. For example:

i) For H_o: diff. = 0 and H_a: diff. ? 0. Type:

display tprob(df,t)

ii) For H_o: diff. 3 0 and H_a: diff. < 0. Type:

display 1 - tprob(df,t) **/ 2**

iii) For H_o: diff. 2 0 and H_a: diff 0. Type:

display tprob(df,t) **/ 2**

Where "diff." is the difference between the two sample means, "df" is the number of the degrees of freedom, and "t" is the value for the t-statistics. Comparing Two Groups — V.1

Questions:

a. Assuming unequal variances between the two samples, test whether men tend to be more confident about their math abilities than women (Don't forget to use the formula for degrees of freedom you learned in this class).

b. State the null and the alternative hypotheses using your own words and mathematical notation (*Be careful*).

c. Based on this test, what can you say about the assumption that men tend to be more confident in math than women?