Lab 6: Some hypothesis tests for the mean of a population

OBJECTIVES: This lab is designed to expand on the ideas of hypothesis testing and t-tests. We begin by addressing the earth measurement data of last lab and then consider one more small data set.

DATA: In 1798, H. Cavendish set out to find out the density of the earth relative to that of water, using a torsion balance. He conducted 29 very precise experiments assessing this relative density. A working assumption which we begin with is that the values of his measurements were taken from a Normal distribution centered on the true relative density. The data from all of his 29 experiments are given below.

5.50	5.55	5.57	5.34	5.42	5.30
5.61	5.36	5.53	5.79	5.47	5.75
4.88	5.29	5.62	5.10	5.63	5.68
5.07	5.58	5.29	5.27	5.34	5.85
5.26	5.65	5.44	5.39	5.46	

DIRECTIONS: Open a new Systat data file and enter the above data into a single column. Follow the instructions below, answering all questions. Write out the answers neatly and turn them in to the lab instructor.

1. We wish to test whether Cavendish's measurements are inconsistent with the modern accepted value of 5.517. To do this, we check to see whether Cavendish's values would cause us to reject the null hypothesis that the modern value is plausible. In this case, the null and alternative hypotheses are

$$\begin{aligned} H_0 : \mu &= 5.517 \\ H_1 : \mu &\neq 5.517. \end{aligned}$$

If the null hypothesis is true, then the quantity

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\bar{x} - 5.517}{s/\sqrt{n}}$$

will have a t-distribution with $n-1$ degrees of freedom. We will be using this quantity as our test statistic. The reason for using this quantity is that we would like to be able to specify the probability distribution for the test statistic when the null hypothesis is true so that we can make quantitative statements about the probability of Type I and Type II errors. A Type I error is when we reject the null hypothesis when in fact it is true. Similarly, a Type II error is when we accept the null hypothesis when in fact the alternative hypothesis is true. Since tests for the mean can be conducted in great generality using t-distributions, these "t-tests" have become quite common. Some of the manuverings below will seem a bit contrived, as SYSTAT is actually not well set up for t-tests. Enter the Cavendish data in a column titled CAVE. Go to the command window and type:

LET TESTMEAN=5.517

Under Stats/Stats/t-test, select CAVE and then TESTMEAN (choosing TESTMEAN and then CAVE corresponds to using $\mu - \bar{x}$ instead of $\bar{x} - \mu$) from the left hand box, and choose "Independent" for the t-test option. This returns the value of the test statistic and the p-value associated with it. This p-value is that associated with a 2-sided test. Think of the absolute observed value of the test statistic (call it abs(T)). The p-value is thus the probability that a random variable taken from a t-distribution with the appropriate degrees of freedom either takes on a value greater than abs(T) or less than -abs(T). This can be thought of as the probability of observing a value more extreme than the one that you actually observed. The result of a t-test is generally said to be "significant" if it allows us to reject the null for a given level of significance α. If the p-value is less than α we can reject the null.

a. What is the value of the test statistic? How many degrees of freedom does the t-distribution have?

b. Can we reject the null at the $\alpha = .05$ level? What does this say about Cavendish's measurements and the current accepted value?

c. What are the critical values for this hypothesis test? SYSTAT can compute these so that you don't have to look them up in tables. Go to the command window and type:

   LET GENSTAT=TIF(α,$n-1$). Here $\alpha = .05$ and $n - 1 = 28$

   This computes the α quantile of a t-distribution with $n-1$ degrees of freedom (There is also a ZIF(α) function for the standard normal).

d. Sketch a t-distribution. Indicate the rejection region and critical values for the above test on this drawing. Also draw an arrow indicating where the observed value fell.

2. DATA II: The second data set that will be used for today's lab is the result of an experiment by Charles Darwin which was performed in the following manner. Pairs of seedlings of the same age, one produced by cross-fertilization and the other by self-fertilization, were grown together so that the members of each pair were reared under nearly identical conditions. The aim was to demonstrate the greater vigour of the cross-fertilized plants. The data are the final heights (in inches) of each plant after a fixed period of time. Darwin consulted Francis Galton about the analysis of these data, and they were discussed further in Ronald Fisher's Design of Experiments.

Cross-Fert.	Self-Fert.
23.5	17.4
12.0	20.4
21.0	20.0
22.0	20.0
19.1	18.4
21.5	18.6
22.1	18.6
20.4	15.3
18.3	16.5
21.6	18.0
23.3	16.3
21.0	18.0
22.1	12.8
23.0	15.5
12.0	18.0

3. The first step we want to take in analyzing this data is a study of the descriptive statistics for looking at the differences in each pair. Create two columns, CROSS and SELF, and fill them in the obvious manner. In the command window, type:

   LET DIFF=CROSS-SELF.

   Calculate and record the mean and standard deviation of DIFF. First, we want to look at the distribution of the differences. One of the options under Graph/Density is "dit". Choose that option to look at DIFF (note: sometimes the "ditplot" is also referred to as a "dotplot"). Based on this plot, what can you say about the distribution of the differences with respect to center, spread, and shape? Does it appear that the cross-fertilized plants are taller than the self-fertilized plants?

4. To attempt to show that cross-fertilization does lead to increased height over self-fertilization, we wish to test the following hypothesis:

$$\begin{aligned} \mu_D &= 0 \\ \mu_D &> 0. \end{aligned}$$

In order to carry out this test, we will treat DIFF as an iid sample of size 15 from a population that is $N(\mu_D, \sigma_D^2)$. Then we perform the hypothesis test described above via:

$$T = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences and s_d is the sample standard deviation of the differences.

a. What is the value of this test statistic?

b. The quantiles below are taken from Table E in your textbook. If we are testing this hypothesis at $\alpha = 0.05$, then what is the critical value for this test? Recall that this is a one-sided test.

	0.90	*0.95*	*0.975*	*0.99*
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.625
15	1.341	1.753	2.131	2.603
16	1.337	1.746	2.120	2.584
17	1.333	1.740	2.110	2.567

c. Sketch the sampling distribution, showing the critical value, rejection region, and the value of the test statistic.

d. What is the p-value and what is its interpretation? To calculate the p-value, go to the command window and type:

LET PVAL=1-TCF(*test statistic*,14). When it asks for *test statistic*, you should enter the value of the test statistic that you found in part a).

e. Do we retain or reject the null hypothesis?

5. Calculate the 2-sided 95% confidence interval for the mean of the differences. The formula is given by

$$\bar{d} \pm t_{\alpha/2,n-1} \frac{s_d}{\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences, s_d is the sample standard deviation of the differences, n is the sample size, and $t_{\alpha/2,n-1}$ is the critical value for a two-sided test. You can find the critical value from the table in question 4b.

If the two samples are identically distributed, then the mean of the differences should be 0. Does 0 lie in the confidence interval? What does this tell you about the null hypothesis?

6. Now calculate the 2-sided 90% confidence interval. Does 0 lie in this confidence interval? Which confidence interval is shorter, the 90% confidence interval or the 95% confidence interval?