# Topic 10. Hypothesis Testing: t-test

Cheng Peng

## Contents



# 1 Introduction

In this note, we discuss the hypothesis test of the means of **normal populations** with the assumption that the population variances are unknown. Unlike the normal test we discussed earlier, the sampling distribution of the sample means is NOT a moral distribution. We will introduce a new distribution to characterize the random behavior of the test statistic.

As an application, we also introduce a special procedure to test the difference between paired samples based on the t-test.

# 2 t-test for Normal Population Means

In the previous topic, we test means of unspecified populations based on large samples using the central limit theorem to derive the normal distribution of the test statistic. When the population is normal and population variance is unknown, then the test statistic

$$TS = \frac{\bar{x} - \mu}{s/\sqrt{n}} \to t_{n-1}$$

We have introduced t-distribution when we constructed confidence intervals for normal population means in an earlier topic. The critical value(s) used to define rejection region(s) are found from the t-table.

Next, we use a numerical example to show the steps for a t-test.

**Example 1.** The yield of alfalfa from a random sample of six test plots is 1.4, 1.6, 0.9, 1.9, 2.2 and 1.2 tons per acre. Assume that the random sample comes from a normal population. Test at the 0.05 level of significance whether this supports the contention that the average yield for this kind of alfalfa is 1.5 tons per acre.

**Solution**: We are given a small raw data set of alfalfa yields $\{1.4, 1.6, 0.9, 1.9, 2.2, 1.2\}$. To perform the test, we calculate the sample statistics from the sample. $n = 6$, $\bar{x} = (1.4 + 1.6 + 0.9 + 1.9 + 2.2 + 1.2)/5 = 1.533$, and $s = 0.472$.

**Step 1**: The claim is clearly specified in the question **" the average yield for this kind of alfalfa is 1.5 tons per acre"**, that is, $\mu = 1.5$.

**Step 2**: The null and alternative hypotheses are given by

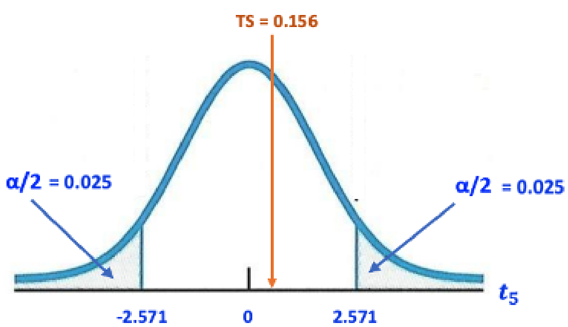
$$H_0 : \ \mu = 1.5 \ \ v.s. \ \ H_a : \ \mu \neq 1.5.$$

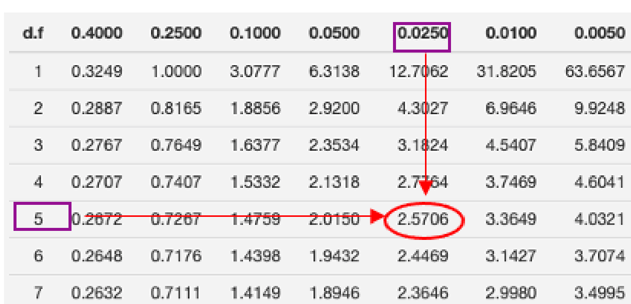
This is a two-tailed test. There are two rejection regions.

**Step 3**: The test statistic is given below

$$TS = \frac{\bar{x} - 1.5}{s/\sqrt{n}} = \frac{1.533 - 1.5}{0.472/\sqrt{6}} \approx 0.171.$$

**Step 4** Since $n = 6$ and the population is normal with an unknown variance. The above test statistic is a t distribution with 5 degrees of freedom. We use the t-table to find the critical values: $CV = \pm t_{0.05/2, 6-1} = \pm t_{0.025, \ \ 5} = \pm 2.571$



| d.f | 0.4000 | 0.2500 | 0.1000 | 0.0500 | 0.0250 | 0.0100 | 0.0050 |
|---|---|---|---|---|---|---|---|
| 1 | 0.3249 | 1.0000 | 3.0777 | 6.3138 | 12.7062 | 31.8205 | 63.6567 |
| 2 | 0.2887 | 0.8165 | 1.8856 | 2.9200 | 4.3027 | 6.9646 | 9.9248 |
| 3 | 0.2767 | 0.7649 | 1.6377 | 2.3534 | 3.1824 | 4.5407 | 5.8409 |
| 4 | 0.2707 | 0.7407 | 1.5332 | 2.1318 | 2.7764 | 3.7469 | 4.6041 |
| 5 | 0.2672 | 0.7267 | 1.4759 | 2.0150 | 2.5706 | 3.3649 | 4.0321 |
| 6 | 0.2648 | 0.7176 | 1.4398 | 1.9432 | 2.4469 | 3.1427 | 3.7074 |
| 7 | 0.2632 | 0.7111 | 1.4149 | 1.8946 | 2.3646 | 2.9980 | 3.4995 |

**Step 5**: Since the test statistic $TS = 0.156$ is NOT in the rejection region, we fail to reject the null hypothesis $H_0 : \ \mu = 1.5$.

**Step 6**: We do not have enough evidence to reject the null hypothesis that the mean yield of the given kind of alfalfa is 1.5 tons. The data tend to support the contention.

**Remarks**: Here are some remarks about the t-test.

1. The steps of the t-test are identical to those in the normal test except for the distribution table used in the test.
2. If the sample size is large, the t-test and the normal test based on the central limit theorem will yield essentially the same result.
3. If the sample size is small and the normal population variance is unknown, we must use the t-test.
4. If the small size is small and the population is NOT normal, we cannot perform any test in this course.

The following video summarizes the above discussions with an example [**17 minutes**].

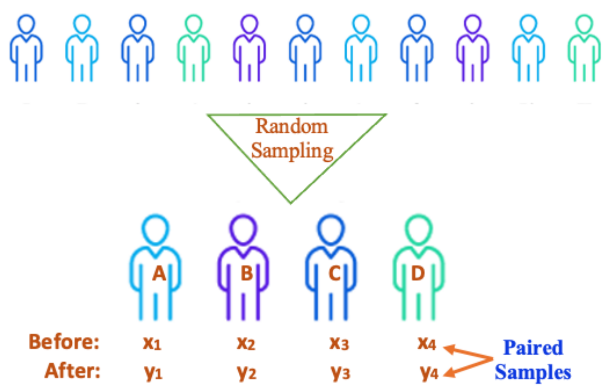
# 3 t-test for Paired Samples

This is a special hypothesis test that involves two samples. The general two-sample tests will be discussed later as a new topic. We can also use the one-sample t-test to generate the solution to the paired t-test.

## 3.1 What Are Paired Samples?

Paired samples are samples taken from the set of subjects under two different conditions such that each observation in one sample can be paired with an observation in the other sample.

To understand the idea of the paired sample, let's consider a hypothetical example. A pharmaceutical manufacturer is developing a new blood pressure drug that requires the FDA's evaluation of safety and effectiveness. To show FDA's review panel the effectiveness of the potential new drug, the clinical trial team recruits a group of subjects following the regulatory requirements. For example, the team selected 4 subjects to participate in the clinical trial. **BEFORE** they receive the new drug, the team took the blood pressure readings from the group of subjects $S_1 = \{x_1, x_2, x_3, x_4\}$ and took another blood pressure readings **AFTER** they received the new drug denoted by $S_2 = \{y_1, y_2, y_3, y_4\}$.




$S_1$ and $S_2$ are called paired samples because $x_1$ and $y_1$, $x_2$ and $y_2$, $x_3$ and $y_3$, $x_4$ and $y_4$ are paired. These paired readings were taken from the same subjects **before** and **after** receiving the treatment.

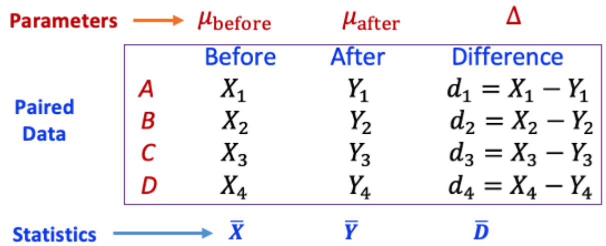
The paired sample method is widely used in many real-world applications.

## 3.2 The Logic of Paired t-test

The object of the paired t-test is to compare the mean measurement between two groups under different conditions in which each observation in one sample can be paired with an observation in the other sample.

If the "before" and "after" means are equal to each other, the drug has no treatment effect. This implies that we assess the treatment effect by comparing the two means of "before" and "after" sample means. For paired samples, we can convert this "two-sample" problem to a one-sample problem and use the regular t-test to compare the two means.

The following figure depicts the structure and notations related to the paired data and the testing procedure.




Recall that our primary interest is test a claim related to $(\mu_{before} - \mu_{after})$ to see the difference between the two means. With the above notation and fact that $\Delta = \mu_{before} - \mu_{after}$, we only need to test claims associated with $\Delta$.

For example, testing the following hypotheses

$$H_0 \ (\mu_{before} - \mu_{after}) = 0 \ \ v.s. \ \ H_a : \ (\mu_{before} - \mu_{after}) \neq 0$$

is equivalent to testing

$$H_0 : \ \Delta = 0 \ \ v.s. \ \ H_a : \ \Delta \neq 0.$$

However, the test based on $\Delta$ only needs to use the single sample data of differences between the paired measurements in the "before" and "after" samples. Therefore, we can use the regular t-test introduced earlier to perform the paired t-test.

## 3.3 Steps for Paired t-test

The steps for paired t-test are the same as those we used before except for the data preparation step added to the 6-step procedure. Due to the nature of the design of the paired sample problems, the size of the sample is usually small. We need to assume the differences of paired measurements, $\{d_1, d_2, d_3, \cdots, d_n\}$, are normally distributed.

In the initial step of data preparation, we need to culculte the mean and standard deviation of the data:

$$D = \frac{d_1 + d_2 + \cdots + d_n}{n}, \quad s_d = \sqrt{\frac{(d_1 - D)^2 + (d_2 - D)^2 + \cdots + (d_n - D)^2}{n-1}}$$

We will use the following example to illustrate the paired t-test.

**Example 2.** To determine whether applying a protective coating to the exterior of a printer increases its operating temperature, 6 printers were selected, and the operating temperature was recorded before and after treatment.

| Printer | no coating | coating |
|---|---|---|
| 1 | 186 | 189 |
| 2 | 185 | 186 |
| 3 | 179 | 183 |
| 4 | 184 | 188 |
| 5 | 183 | 181 |
| 6 | 186 | 188 |

Assuming that the temperatures are normally distributed. Does the data support the theory that the coating increases the mean operating temperature? $\alpha = 0.05$.

**Solution**: We first calculate the mean and standard deviation of the differences (the last column in the following table).

| No-coating | coating | Difference |
|---|---|---|
| 186 | 189 | 3 |
| 185 | 186 | 1 |
| 179 | 183 | 4 |
| 184 | 188 | 4 |
| 183 | 181 | -2 |
| 186 | 188 | 2 |

After some algebra (using the given formulas), we have $D = 2$ and $s_d = 2.28$.

**Step 1**. The claim is that *coating increases the mean operating temperature*. That is D = coating. temp – no-coating.temp > 0.

**Step 2**. Setting up the null and alternative hypotheses

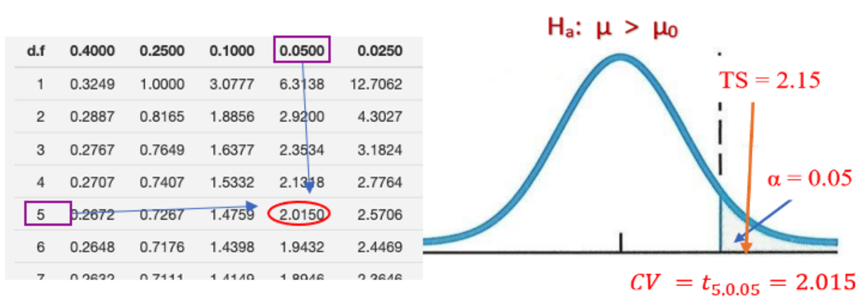
$$H_0 : \ \Delta \leq 0 \ \ versus \ \ H_a : \ \Delta > 0$$

This is a right-tailed test.

**Step 3**. The test statistic is defined by

$$TS = \frac{D - 0}{s_d/\sqrt{n}} = \frac{2 - 0}{2.28/\sqrt{6}} \approx 2.15.$$

**Step 4**. The critical value of this right-tailed test is given by $CV = t_{5,0.05} = 2.015$. This means that the rejection region is $RR = (2.015, \infty)$. All information is summarized in the following figure.



| d.f | 0.4000 | 0.2500 | 0.1000 | 0.0500 | 0.0250 |
|---|---|---|---|---|---|
| 1 | 0.3249 | 1.0000 | 3.0777 | 6.3138 | 12.7062 |
| 2 | 0.2887 | 0.8165 | 1.8856 | 2.9200 | 4.3027 |
| 3 | 0.2767 | 0.7649 | 1.6377 | 2.3534 | 3.1824 |
| 4 | 0.2707 | 0.7407 | 1.5332 | 2.1318 | 2.7764 |
| 5 | 0.2672 | 0.7267 | 1.4759 | 2.0150 | 2.5706 |
| 6 | 0.2648 | 0.7176 | 1.4398 | 1.9432 | 2.4469 |



**Step 5**. Since the test statistic is in the rejection region, we **reject** the null hypothesis $H_0 : \ \Delta \leq 0$ and **conclude** the alternative hypothesis $H_a : \ \Delta > 0$.

**Step 6**. We conclude that the protective coating on the exterior of a printer increases its operating temperature.

**A Cautionary Remark**. When calculating the differences of paired measurements, we take the form of either *before - after* or *after-before*. However, the form of the difference will impact the form of the claim. In the above example, the claim is **coating increases the mean operating temperature**, if $d$ = *coating - no.coating*, the form of the claim is $\Delta > 0$; if $d$ = *no.coating - coating*, the form of the claim is $\Delta < 0$.

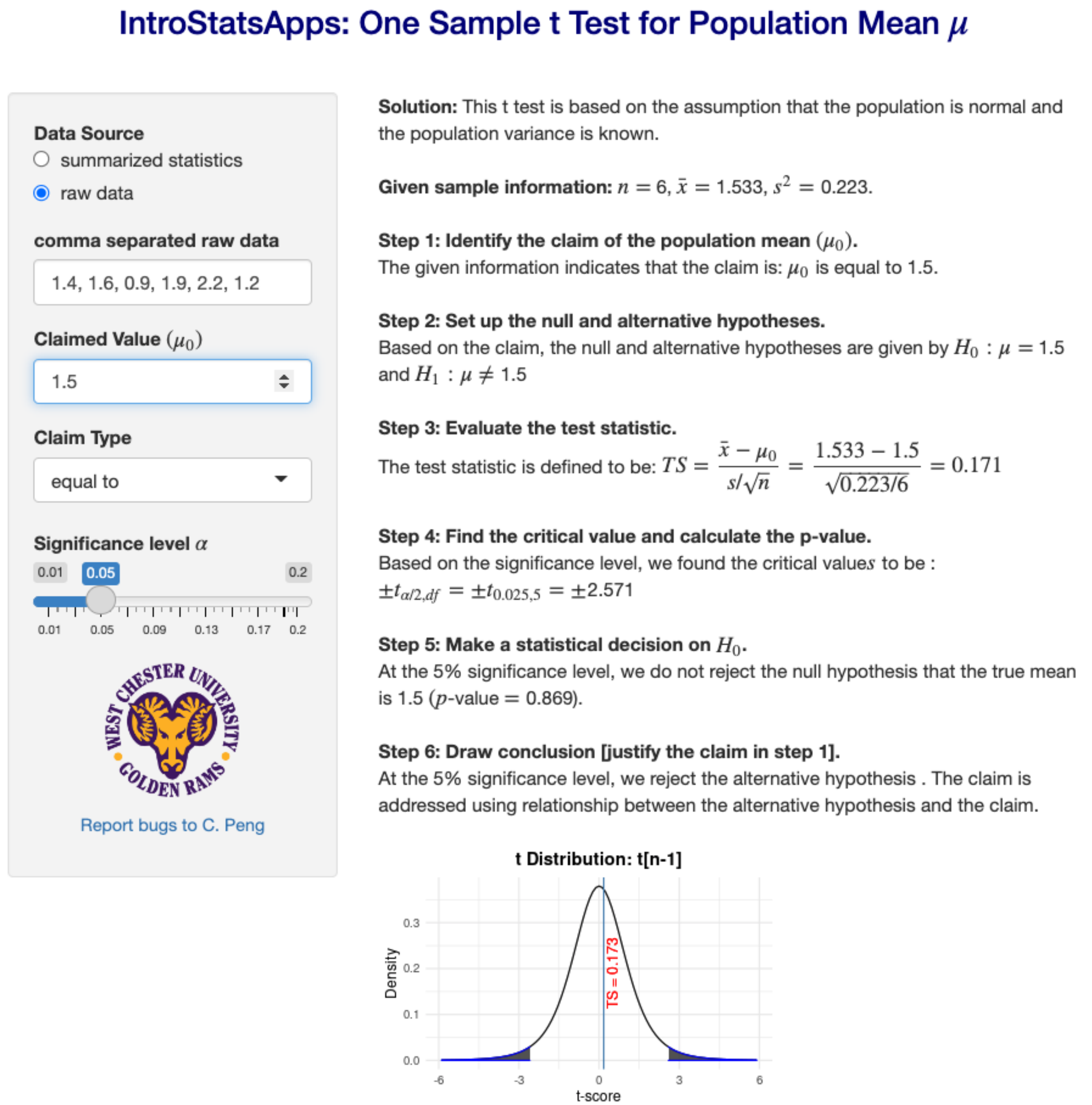
The following video summarizes the above discussions with an example [**20 minutes**].

# 4 Use of Technology

Stats App *one sample t-test* is available at: (https://wcu-peng.shinyapps.io/oneMean-ttest/). You can use this app to check your work.

## 4.1 One-sample t-test

We use the Apps to generate the solution of the above **example 1**. Since we have raw data (i.e., individual data values), we need to type in all values with adjacent values separated by a comma.

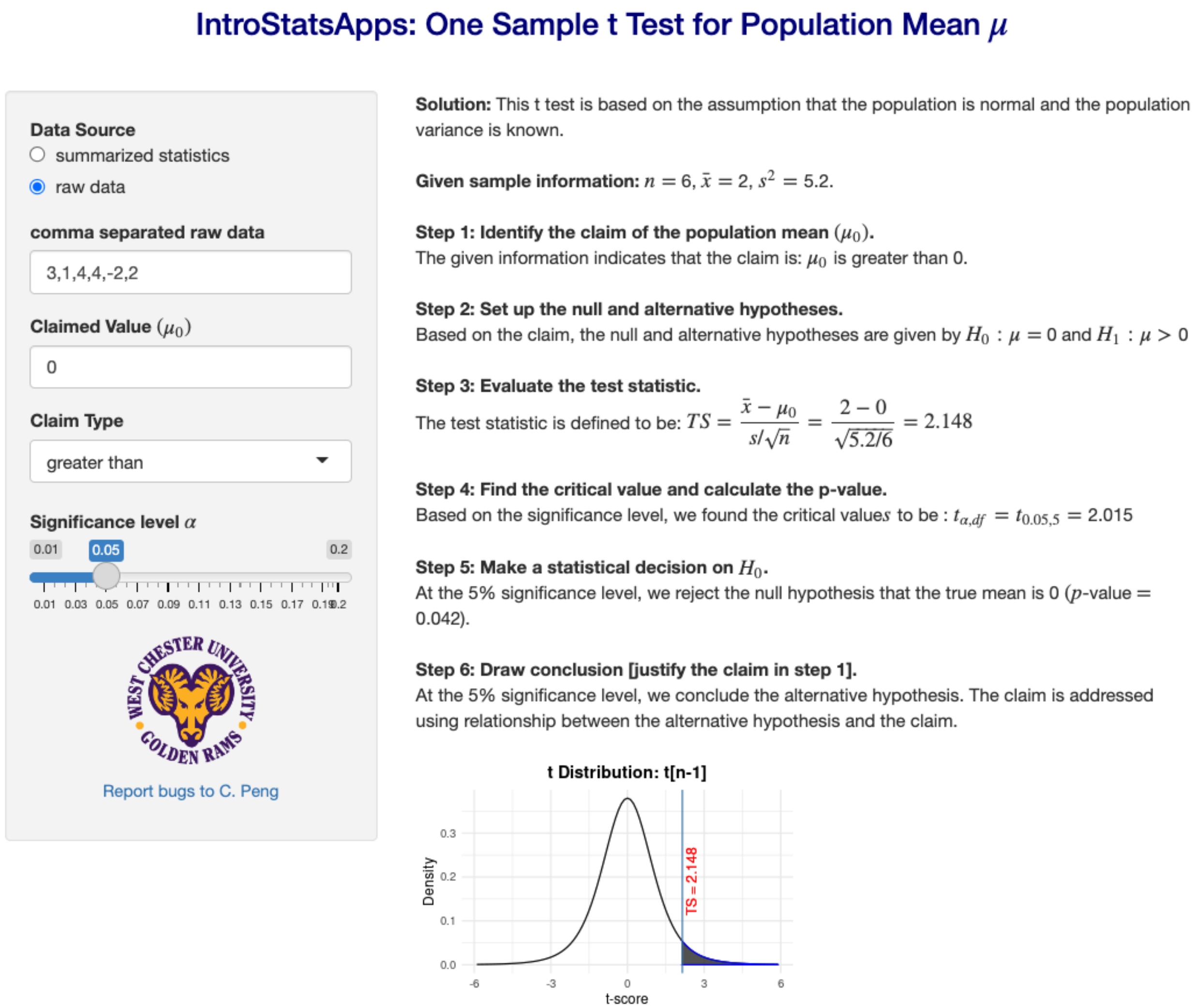



The generated solution is essentially the same as the manual solution except for small rounding-up errors.

## 4.2 Paired t-test

Before using this for paired t-test, we first calculate the difference. Then based on the form of the difference specified in the claim and provide this information to the app. To illustrate this application, we generate the solution of the above example 2.

Recall the claim and the form of difference used in example 2: D = coating. temp − no-coating.temp > 0. The set of differences is $\{3, 1, 4, 4, -2, 2\}$ and will be typed in the raw data input box.




# 5 Practice Exercises

Practice the following exercises and use the app to check your work.

1. We have the potato yield from 12 different farms. We know that the standard potato yield for the given variety is $\mu = 20$. Test if the potato yield from these farms is significantly better than the standard yield using the following random sample.

   21.5, 24.5, 18.5, 17.2, 14.5, 23.2, 22.1, 20.5, 19.4, 18.1, 24.1, 18.5

2. Two different tips are available for a hardness-testing machine. The machine operates by pressing the tip into a metal specimen and then measuring the depth of the resulting depression. Eight metal specimens are chosen, and each specimen is tested with both tips. Assuming that depths are normally distributed, and the resulting depths are shown below (coded). At level $\alpha = 0.05$, is there any difference between the two tips?

| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |
|---|---|---|---|---|---|---|---|---|
| Tip 1 | 16 | 3 | 10 | 12 | 4 | 7 | 3 | 9 |
| Tip 2 | 13 | 3 | 9 | 9 | 4 | 4 | 4 | 5 |

3. A golf club manufacturer claims that golfers can lower their scores by using the manufacturer's newly designed golf clubs. Eight golfers are randomly selected, and each is asked to give his or her most recent score. After using the new clubs for one month, the golfers are again asked to give their most recent score. The scores for each golfer are shown in the table. Assuming the golf scores are normally distributed, is there enough evidence to support the manufacturer's claim at $\alpha = 0.10$?

| Golfer | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Score (old) | 89 | 84 | 96 | 82 | 74 | 92 | 85 | 91 |
| Score (new) | 83 | 83 | 92 | 84 | 76 | 91 | 80 | 91 |

4. To assess whether or not a certain training program can increase the max vertical jump (in inches) of college basketball players, we recruit a simple random sample of 20 college basketball players and measure each of their max vertical jumps and then have each player use the training program for one month. At the end of the month, we measure their max vertical jump again. The following table records the measurements.

| Player | Max Vertical Jump Before Training Program | Max Vertical Jump After Training Program |
|---|---|---|
| Player 1 | 22 | 24 |
| Player 2 | 20 | 22 |
| Player 3 | 19 | 19 |
| Player 4 | 24 | 22 |
| Player 5 | 25 | 28 |
| Player 6 | 25 | 26 |
| Player 7 | 28 | 28 |
| Player 8 | 22 | 24 |
| Player 9 | 30 | 30 |
| Player 10 | 27 | 29 |
| Player 11 | 24 | 25 |
| Player 12 | 18 | 20 |
| Player 13 | 16 | 17 |
| Player 14 | 19 | 18 |
| Player 15 | 19 | 18 |
| Player 16 | 28 | 28 |
| Player 17 | 24 | 26 |
| Player 18 | 25 | 27 |
| Player 19 | 25 | 27 |
| Player 20 | 23 | 24 |

5. A professor wants to know whether the average scores of quiz 1 and quiz 2 are different. The scores of the two quizzes are given below.

| Student ID | Quiz 1 | Quiz 2 |
|---|---|---|
| 001 | 98 | 94 |
| 002 | 100 | 98 |
| 003 | 95 | 98 |
| 004 | 90 | 88 |
| 005 | 90 | 89 |
| 006 | 92 | 91 |
| 007 | 80 | 84 |
| 008 | 78 | 80 |
| 009 | 88 | 88 |