t-Intervals for Two Independent Samples

1. For each of the following, state whether you'd use one-sample or two-sample procedures to analyze the data, and explain why.

    1. A university is interested in whether there's a difference between students who live on campus and students who live off campus with respect to absenteeism. Over one semester, researchers take random samples of on-campus and off-campus students and record the number of classes each student misses.

        You should use a two-sample procedure to analyze the data since you're analyzing two different samples; one population of students who live on campus and another population of students who live off campus.

    2. A company wants to know whether listening to music improves efficiency among its employees. Two employees are randomly chosen from each of 10 department and are given the same task to complete. One employee from each pair is allowed to listen to music and the other employee isn't. Researchers record the time it takes each employee to complete the assigned task.

        You should use a one-sample test because there is natural pairing involved with the creation of both populations. A pair is made with one employee allowed to listen to music and another who is not.

    3. A restaurant owner wants to know whether customers will drink more coffee if he switches to a new brand. Every night for one week, he randomly selects one half of the restaurant and serves the new brand to customers who sit in this section. Customers in the other half receive the old brand. He records the number of refills ordered in each section.

        You should use a two-sample test because we are trying to find the difference between two different populations: one population which receives the new type of coffee and another which receives the older brand of coffee.

    4. You want to know if there's a difference between the time it takes you to get to school on the bus and the time it takes you to get to school when you get a ride from a friend. Over a semester, you flip a coin to decide whether you'll take the bus or get a ride. You record your travel time daily.

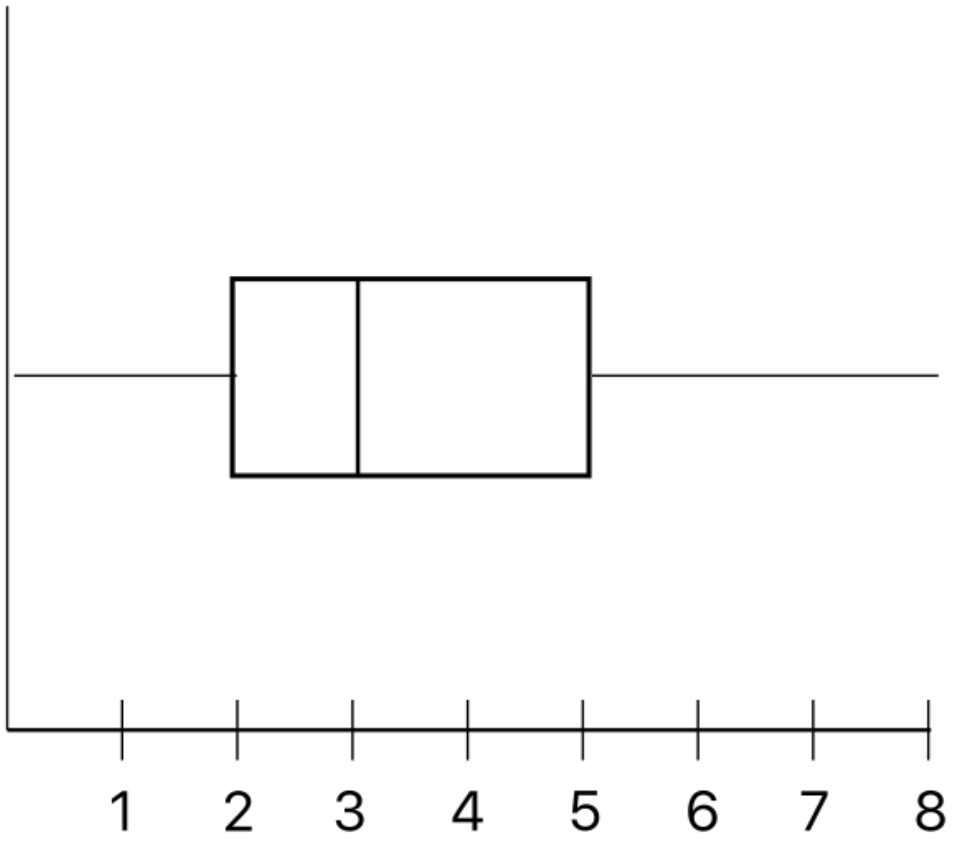
We would use a two-sample test because we are looking at the difference between two independent populations; the travel time when you take the bus, and the travel time when you get a ride from friend

2. A university is interested in whether there's a difference between students who live on campus and students who live off campus with respect to absenteeism. Over one semester, researchers take random samples of on-campus and off-campus students and record the following number of missed classes over a semester...

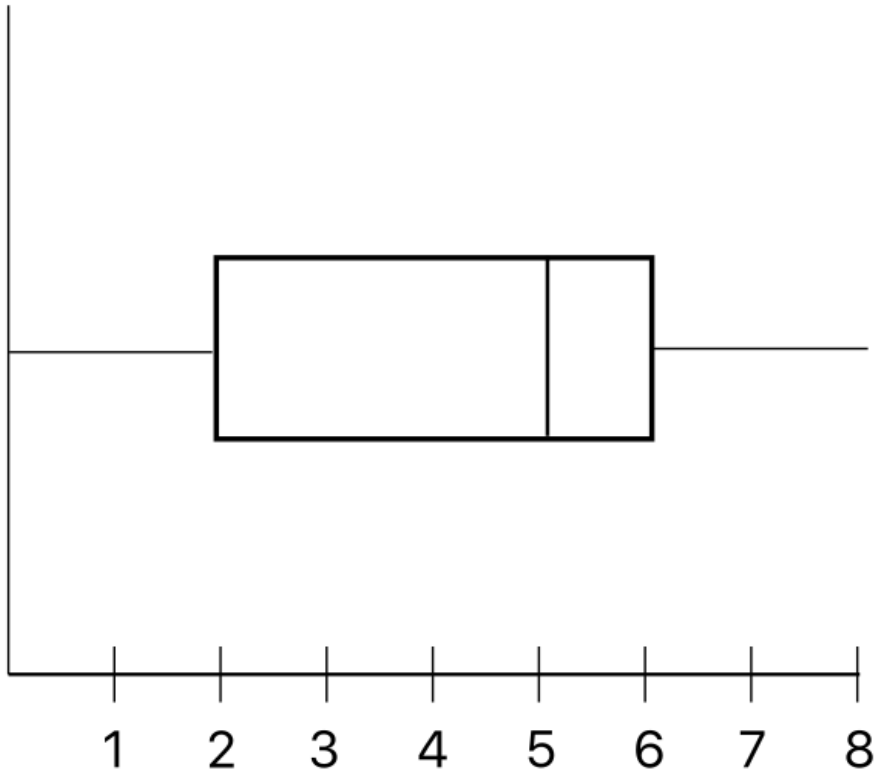
    1. Would you use a t confidence interval or a z confidence interval to determine whether there's a significant difference between the two groups? What are the conditions for using this kind of confidence interval? Do these data meet the necessary conditions? Use sketches of modified box-and-whisker plots to support your decision.

        I would use a t confidence interval. The conditions to use a t confidence interval are that it must be an SRS and that the population must be normally distributed (however, since t is much more robust than z, it's fine if the population is a little skewed)

        Off-Campus



        On-Campus



As observed by the box plots, there is some skew in both sample populations; however, using the t distribution is optimal since it is more robust.

2. What are the degrees of freedom for this test using the conservative method?

   The conservative method is best, so we have to use the smallest degree of freedom between the two samples. Since both samples have 15 points, the degrees of freedom is 14.

3. What are the sample statistics for this test? Consider on-campus students to be sample one and off-campus students to be sample two.

   Sample 1 Mean: 3.33

   Sample 1 SD: 1.7593

   Sample 1 Sample Size: 15

   Sample 2 Mean: 4.13

   Sample 2 SD: 2.53

   Sample 2 Sample Size: 15

   t-value: -1.01

4. Compute a 95% confidence interval for the difference between the number of classes missed by each group of students.

$$2.145 \times \sqrt{\frac{1.7593^2}{15} + \frac{2.5317^2}{15}} = 1.7075$$

$$\boxed{\text{-0.8} \pm 1.7075}$$

5. Compute a 90% confidence interval for the difference between the number of classes missed by each group of students.

$$1.761 \times \sqrt{\frac{1.7593^2}{15} + \frac{2.5317^2}{15}} = 1.4018$$

$$\boxed{\text{-0.8} \pm 1.4018}$$

6. Based on the two confidence intervals you computed in parts d and e, draw a conclusion about the differences between the means of the two groups. Use $H_0 : (\mu_1 - \mu_2) = 0$ as your null.

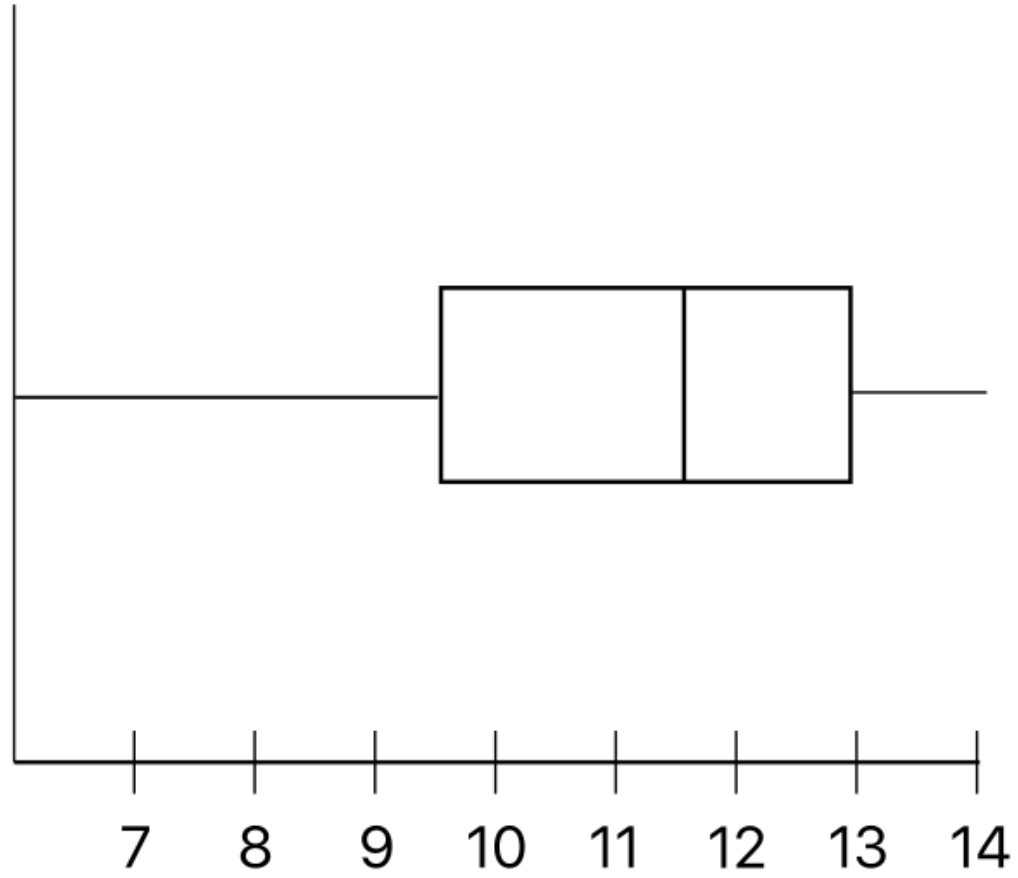
   The null hypothesis is between both the confidence intervals. Therefore, there is no statistically significant difference between the amount of absences between off-campus and on-campus students.

3. Over the past semester, you've collected the following data on the times it takes you to get to school by bus and by car...You want to know if there's a difference in the time it takes you to get to school by bus and by car.

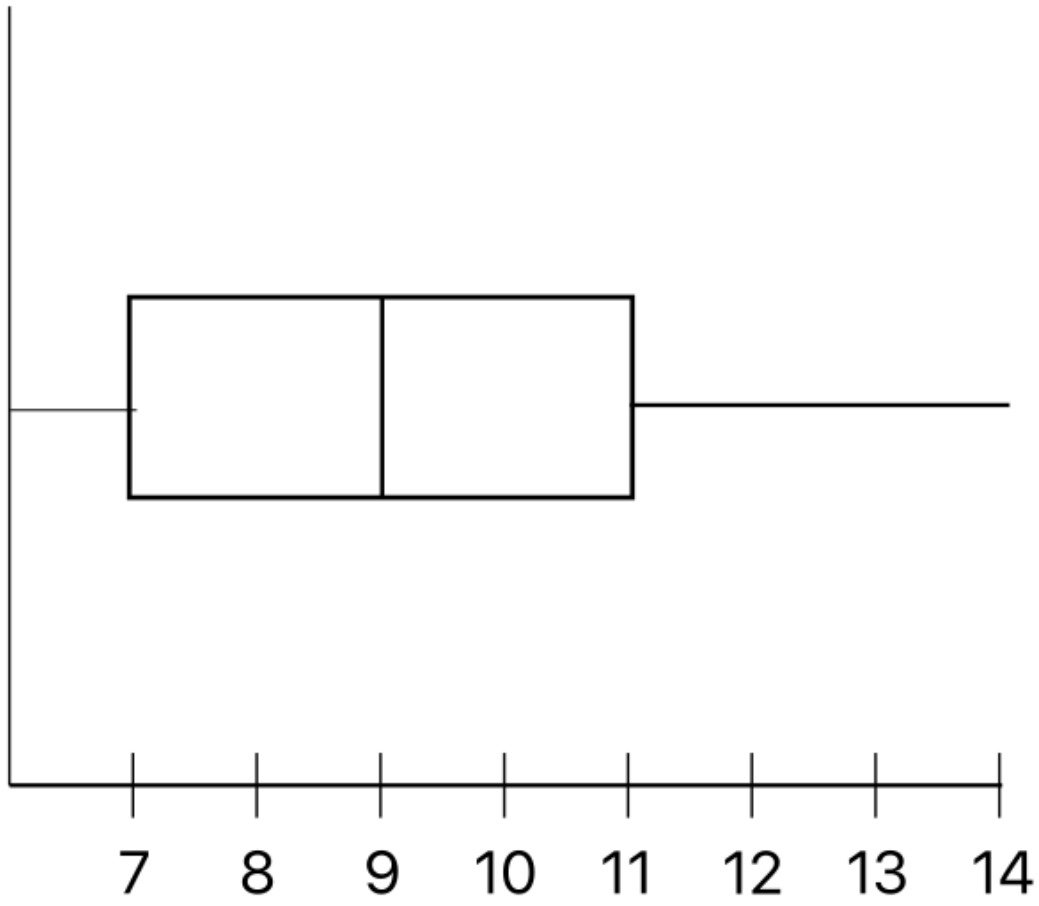
   1. What test would you use to look for a difference in the two data sets, and what are the conditions for this test? Do the data meet these conditions? Use sketches of modified box-and-whisker plots to support your decision.

      We would use a t-confidence interval. To use this, the sample must be an SRS and it must come from a normally distributed population.

      Bus



Car



This data looks nearly symmetrical and very "normal". Therefore, the t-distribution is appropriate.

2. What are the degrees of freedom (k) for this test using the conservative method?

   The conservative method depends on finding the smaller degree of freedom between the two samples. Sample 1 has 20 points and Sample 2 has 18, so we use Sample 18 to calculate k. 18-1 = 17.

3. What are the sample statistics for this test? Consider the data you collected for bus times to be sample one and the data for car times to be sample two.

   Sample 1 Mean: 11.3

   Sample 1 SD: 2.387

   Sample 1 Size: 20

   Sample 2 Mean: 9.06

   Sample 2 SD: 2.287

   Sample 2 Size: 18

   t-value: 2.959

4. Compute a 99% confidence interval for the difference between the time it takes you to get to school on the bus and the time it takes to go by car. Draw a conclusion about this difference based on this confidence interval using $H_0 : (\mu - \mu_2) = 0$ as your null.

$$2.878 \times \sqrt{\frac{2.3864^2}{20} + \frac{2.2874^2}{18}} = 2.1832$$

$$\boxed{2.24 \pm 2.1832}$$

5. Construct the same confidence interval you did in part d, this time using your graphing calculator. show what you do on your calculator and what you put into your calculator, and give the confidence interval and degrees of freedom.

   MENU,6,6,4

   select data sets

$$\boxed{2.24 \pm 2.0634}$$

   df = 35.844

6. How is the interval computed on a calculator different from the interval computed hand? Why is it different? In this case, would you come to a different conclusion for the hypothesis $H_0 : (\mu - \mu_2) = 0$ if you used the confidence interval generated by the calculator?

The calculator does not use the conservative method, increasing the degrees of freedom. This causes a more normal distribution and which has a smaller chance of error. We do not change our conclusion; it is outside the confidence interval calculated by the calculator as well.