

Unit 27: Comparing Two Means



SUMMARY OF VIDEO

It's an age old battle of the sexes. Are men or women worse drivers? Whatever your opinion on this question, a statistician needs evidence in order to make a decision. One way to analyze this question would be to see which gender, on average, gets more moving violations. We could take a sample from all licensed drivers in one state, and then look at the number of tickets each person received in one year. We could then calculate the mean number of tickets received by members of each gender and compare the two numbers to see which group had the worst driving record.

Comparing two populations is an important topic in statistics because it occurs quite frequently. Moreover, we can use inference to move beyond just looking at two sample means, as was suggested in our driving example. We can go on to figure out whether the difference between two groups is statistically significant; and if it is, we can calculate a confidence interval for the difference of population means.

That's what researchers did when they decided to investigate the difference in the amount of calories necessary to power daily life in two groups of people with very different lifestyles. Herman Pontzer is an anthropologist who is interested in how energy is used by primate species, particularly human beings. Pontzer teamed up with other researchers to work with the Hadza in Tanzania, a group of traditional hunter-gatherers who live in a way very similar to our ancestors. Men hunt with bows and arrows and women forage for plant foods and dig for root vegetables. The Hadza are a lot more active and cover a lot more ground than their Western counterparts. Everyone had always assumed that this physically-demanding hunter-forager lifestyle would require much more energy than the relatively inactive daily life of a Western office worker. In fact, one suspected cause of the obesity epidemic in the West is our more sedentary modern lifestyle. But the Hadza's actual energy expenditure had never yet been tested.

Was the assumption correct that the Hadza used more calories throughout their day? Pontzer and his team already had data on how many calories typical Americans and Europeans burned in their daily lives. Now they needed to measure how many calories it took to power

the Hadza through their daily tasks. They used a technique that relies on the subjects drinking something called doubly-labeled water. For this technique, a person drinks some water where the hydrogen and oxygen have been enriched with rare isotopes. From urine samples taken over a two-week period, researchers can measure with the use of spectroscopy how much of those rare isotopes are in their urine samples. As the concentration of the special traceable hydrogen and oxygen isotopes in the urine goes down over time, Pontzer can figure out how much carbon dioxide the subject has exhaled over the course of the study. When a body burns calories, a byproduct is carbon dioxide, so the amount of carbon dioxide exhaled told the researchers how much energy the Hadza were expending. In addition, the researchers recorded the physical activity of the Hadza by having them wear heart rate monitors and GPS units.

The Hadza are typically smaller and lighter than their Western counterparts. That difference required Pontzer and his colleagues to use sophisticated statistical techniques in their analyses to control for the effects of body size, age, and sex. To keep things simple, so that we can follow their comparison, we'll look just at women with comparable body sizes from the Hadza and Western groups. We want to use our sample to determine whether there is a significant difference between the means of the Hadza and Western populations.

First, the scientists calculated the mean total energy expenditure (TEE), which was measured in calories, for each group. The sample means, standard deviations, and sample sizes for each group are as follows:

Hadza

$$\bar{x}_1 = 1,877 \text{ calories, } s_1 = 364 \text{ calories, } n_1 = 17$$

Westerners

$$\bar{x}_2 = 1,975 \text{ calories, } s_2 = 286 \text{ calories, } n_2 = 26$$

Is the difference between these sample means significant? Or, could the difference we see be due simply to chance variation? We can set up a significance test to figure this out. Below are the null and alternative hypotheses concerning the total energy expenditure.

H_0 : Mean Hadza TEE = Mean Western TEE

$$\mu_1 = \mu_2$$

H_a : Mean Hadza TEE \neq Mean Western TEE

$$\mu_1 \neq \mu_2$$

The two-sample t -test statistic is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Now we can substitute the numbers into the formula. We have the sample means, standard deviations, and sample sizes. For the value of $\mu_1 - \mu_2$, we use the value from the null hypothesis, which states these two means are equal, and hence, $\mu_1 - \mu_2 = 0$.

$$t = \frac{(1,877 - 1,975) - 0}{\sqrt{\frac{(364)^2}{17} + \frac{(286)^2}{26}}} \approx -0.94$$

Like all of the z - or t -test statistics that we have encountered, this one tells us how far $\bar{x}_1 - \bar{x}_2$ is from 0, the hypothesized difference in means, in standard units.

Software can figure out the degrees of freedom for the t -test statistic, or we can just go with a very conservative approach that uses the smaller sample size minus one, which gives us 16 degrees of freedom. We can look up the corresponding p -value in a t -table or use technology; either way, we get $p = 0.3612$. That means that assuming the null hypothesis is true, we have a 36% chance of seeing a t -value as or more extreme than the one we calculated. A 36% chance is pretty likely, so we have insufficient evidence to reject the null hypothesis. We conclude that there is no significant difference between total energy expenditure of Hadza women and Western women.

This, in fact, is what the researchers concluded. After controlling for body size, age, and sex, the scientists did not find any statistical difference when they compared the mean daily energy expenditure of the Hadza and the Westerners. This result seemed counterintuitive, since they knew the Hadza were much more active. The researchers suspect that the Hadza's bodies are allocating a smaller percentage of those daily calories to run-of-the-mill cellular function and more to physical activity. Researchers think that it is a difference in energy allocation, not energy efficiency.

Today's obesity epidemic tells us something is out of balance between the amount of calories that we take in and the amount we burn off. Based on this study and others, metabolism seems to hold quite constant among different populations of people with varying activity levels. Because of this finding, Pontzer and his colleagues place the blame for rising societal levels of obesity more on people eating too much than on our modern lifestyle.

STUDENT LEARNING OBJECTIVES

- A. Understand when to use two-sample t -procedures.
- B. Know how to check whether the underlying assumptions for a two-sample t -procedure are reasonably satisfied.
- C. Be able to calculate a confidence interval for the difference of two population means.
- D. Be able to test hypotheses about the difference between two population means. Be able to calculate the t -test statistic and use technology to determine a p -value.

CONTENT OVERVIEW

Consider the following questions: Do men earn more than women? Are women smarter than men? Do students from private schools do better on their SATs than students from public schools? Which relieves headaches more quickly, Tylenol or Advil? Each of these questions seeks to compare two populations or two treatments – a commonly encountered situation in statistical practice.

We begin with a question related to the activity in Unit 26: Are the step lengths of 10th-grade male students longer, on average, than the step lengths of 10th-grade female students? In this case, the comparison is between two populations, 10th-grade males and 10th-grade females. Let μ_1 and σ_1 be the mean and standard deviation, respectively, of step lengths for the population of 10th-grade males. Let μ_2 and σ_2 be the mean and standard deviation, respectively, for the 10th-grade females. If there is no difference between the mean step lengths of male students and female students, then $\mu_1 - \mu_2 = 0$. However, if males, on average, take longer steps than females, then $\mu_1 - \mu_2 > 0$. We can state the null hypothesis and alternative hypothesis as follows:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 > 0$$

Suppose we randomly select two samples, one of size n_1 from the male students and another of size n_2 from the female students. After collecting the data, we can calculate the sample means, \bar{x}_1 from the males and \bar{x}_2 from the females, and sample standard deviations, s_1 from the males and s_2 from the females. It seems reasonable to use the difference in sample means, $\bar{x}_1 - \bar{x}_2$, to estimate the difference in population means, $\mu_1 - \mu_2$. If the two populations are normally distributed or if the sample sizes are large, then $\bar{x}_1 - \bar{x}_2$ has a normal distribution with the following mean and standard deviation:

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 \text{ and } \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Transforming $\bar{x}_1 - \bar{x}_2$ into a z-score, we get:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The two-sample z-test statistic has the standard normal distribution. At this point, if we knew the population standard deviations, we could use a z-procedure to test our hypotheses. Unfortunately, σ_1 and σ_2 are unknown, so we will need to use the sample standard deviations, s_1 and s_2 , as estimates. Substituting s_1 and s_2 in place of σ_1 and σ_2 gives us the **two-sample t-test statistic**:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The two-sample t-test statistic has an approximate t-distribution. The degrees of freedom (*df*) are a bit complicated to figure out. We can either use software or adopt a conservative approach and set the degrees of freedom to be one less than the smaller of the two sample sizes.

Now, we return to our hypotheses about step lengths of male and female students. Sample data were collected and are summarized below:

Males: $n_1 = 12$, $\bar{x}_1 = 64.08$ cm, $s_1 = 7.71$ cm

Females: $n_2 = 15$, $\bar{x}_2 = 60.34$ cm, $s_2 = 7.74$ cm

Normal quantile plots of the male and female step-length data indicate that it is reasonable to assume that step lengths are approximately normally distributed. Now we are ready to compute the t-test statistic. Using the null hypothesis value of 0 for $\mu_1 - \mu_2$ and our sample means and standard deviations, we get:

$$t = \frac{(64.08 - 60.34) - 0}{\sqrt{\frac{(7.71)^2}{12} + \frac{(7.74)^2}{15}}} \approx 1.25$$

Taking the conservative approach, we assume $df = 12 - 1$, or 11. Since the alternative is one-sided, we use technology to determine the area under the t -density curve that lies to the right of 1.25, our observed value of the test statistic. As shown in Figure 27.1, this gives $p = 0.1186$.

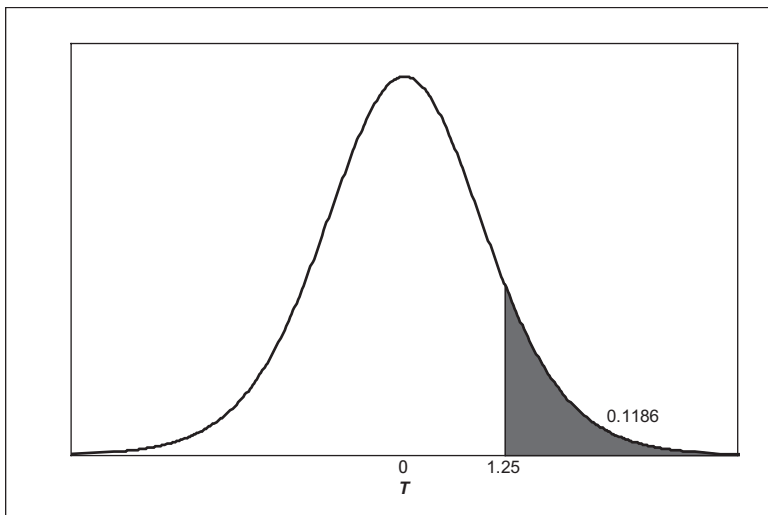


Figure 27.1. Density curve for t -distribution with $df = 11$.

Since $p > 0.05$, there is insufficient evidence to reject the null hypothesis. We cannot conclude that the mean step length of 10th-grade male students differs from the mean step length of 10th-grade female students.

The next example involves a study that compares two teaching strategies for nursing students – lecture notes combined with structured group discussions versus a traditional lecture format. Two groups of students taking a medical-surgical nursing course were taught using each of the two strategies. Exam scores were used to compare the effectiveness of the two teaching strategies. Let μ_1 be the mean exam score for students enrolled in the lecture notes/group discussion version of the course; let μ_2 be the mean exam score for students enrolled in the lecture only version of the course. We set up the following null and alternative hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

Exam scores of two groups of students taught by each of these methods were collected with the following results:

Lecture notes/group discussion: $n_1 = 81$, $\bar{x}_1 = 80.6$, $s_1 = 7.34$

Lecture format (control): $n_2 = 88$, $\bar{x}_2 = 77.68$, $s_2 = 7.23$

Since the sample sizes are large, we can conduct a t -test to decide between the null and alternative hypotheses without first checking that the data come from normal distributions.

Here are the calculations:

$$t = \frac{(80.60 - 77.68) - 0}{\sqrt{\frac{7.34^2}{81} + \frac{7.23^2}{88}}} \approx 2.60$$

Adopting the conservative approach, we use $df = 81 - 1$, or 80, to determine the p -value. Based on Figure 27.2, we get a value of $p = 2(0.00555) \approx 0.011$. (Note: In this situation, we have a two sided alternative.)

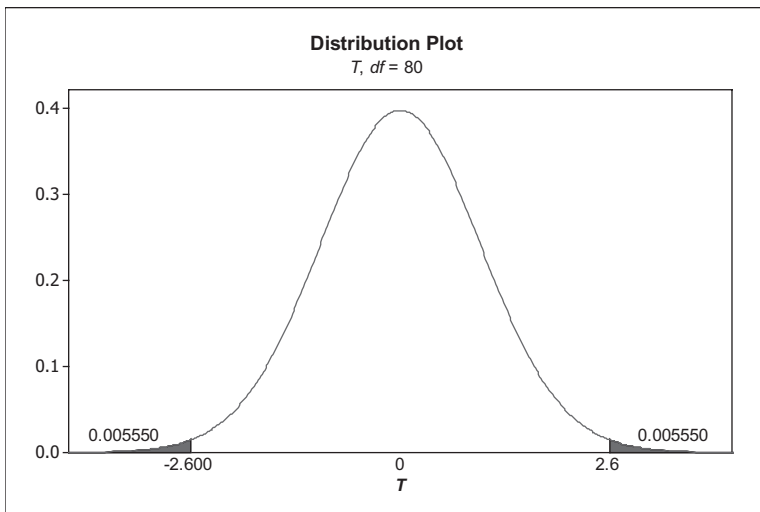


Figure 27.2. Calculating the p -value for a two-sided alternative.

Since $p < 0.05$, we reject the null hypothesis and accept the alternative hypothesis that the mean exam scores for the two teaching methods differ. To estimate that difference, we calculate a two-sample t -confidence interval for $\mu_1 - \mu_2$ using the following formula:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Adopting the conservative approach, we set $df = 80$ and determine a t -critical value for a 95% confidence level: $t^* = 1.990$. Now we are ready to calculate a 95% confidence interval for $\mu_1 - \mu_2$:

$$(80.60 - 77.68) \pm (1.990) \sqrt{\frac{7.34^2}{81} + \frac{7.23^2}{88}} \approx 2.92 \pm 2.23, \text{ or } (0.69, 5.15).$$

Hence, the mean exam scores for the lecture notes/group discussion teaching strategy are between 0.69 and 5.15 points higher than the mean exam scores for the traditional lecture teaching strategy.

We conclude this section with one final comment related to checking the underlying assumptions for the **two-sample t -procedures**. In the development of the two-sample t -procedures for cases where the sample sizes are small, we assume that the population distributions are normal. As it turns out, if the sample sizes are reasonably close and the population distributions are similar in shape, without major outliers, the probabilities from the t -distribution are quite accurate even if the population distributions are not normal.

KEY TERMS

Two sample t -procedures are used to test or estimate $\mu_1 - \mu_2$, the difference in the means of two populations (or treatments). The required data consists of two independent simple random samples of sizes n_1 and n_2 from each of the populations (or treatments).

The **two-sample t -test statistic** for testing the difference in population means is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where the value for $\mu_1 - \mu_2$ is taken from the null hypothesis. There are two options for finding the degrees of freedom (df) associated with t : (1) use technology or (2) use a conservative approach and let $df = \text{smaller of } n_1 - 1 \text{ or } n_2 - 1$.

The **two-sample t -confidence interval** for $\mu_1 - \mu_2$ is computed as follows:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The degrees of freedom for calculating t^* , the t -critical value associated with the confidence level, uses the approach outlined for the two-sample t -test statistic.

THE VIDEO

Take out a piece of paper and be ready to write down answers to these questions as you watch the video.

1. How might a statistician gather evidence to answer the following question: Are men or women worse drivers?
2. What was different about the lifestyle of the Hadzas compared to typical Europeans or Americans?
3. What was Pontzer's original assumption about the daily energy expenditure (in calories) consumed by the Hadza compared to the Westerners?
4. What type of test was used in the video to test this assumption for Hadza and Western women of similar body size?
5. Was the assumption in question 3 correct? Explain.
6. On what did Pontzer and his colleagues place the blame for rising societal levels of obesity?

UNIT ACTIVITY:

CHIPS AHOY, REGULAR AND REDUCED FAT

Nabisco's Chips Ahoy is a popular brand of chocolate chip cookies. Nabisco makes both a regular and, for those who want to restrict their fat intake, a reduced fat version of chocolate chip cookies. The question for this activity is to find out whether the mean number of chips per cookie is the same for Chips Ahoy reduced fat chocolate chip cookies as it is for Chips Ahoy regular chocolate chip cookies.



1. If needed, collect the data on the number of chips per cookie in regular and reduced fat Chips Ahoy cookies. Your instructor will provide directions. (You may already have collected the data as part of Unit 25's activity.)
2. a. Do you think the mean number of chips per cookie is the same for both Chips Ahoy regular and Chips Ahoy reduced fat chocolate chip cookies? If not, which type, regular or reduced fat, do you think has, on average, more chips per cookie? Explain.
b. Set up null and alternative hypotheses for testing whether the mean number of chips per cookie is the same for both the regular and the reduced fat version of Chips Ahoy chocolate chip cookies. Be sure to define any symbols that you use in your hypotheses. (Did you choose a one-sided or two-sided alternative?)
3. Report the sample size, mean and standard deviation for the regular chocolate chip cookie data. Then do the same for the reduced fat chocolate chip cookie data.
4. Make comparative graphic displays of the chip count data for the regular and reduced fat cookies. Based on your plots, do the chip counts for the two types of cookies appear to differ?

5. a. Compute the two-sample t -test statistic. Show your calculations.
- b. Determine a p -value for your test statistic in (a).
- c. Is there a significant difference in the mean number of chips per cookie in regular and reduced fat Chips Ahoy chocolate chip cookies? Explain.
6. Calculate a 95% confidence interval for the difference between the mean number of chips per cookie in Chips Ahoy regular and Chips Ahoy reduced fat chocolate chip cookies.

EXERCISES

1. A study published in the *Journal of Business and Psychology* investigated whether being pregnant had an adverse effect on women's job performance appraisal ratings. Two groups of female employees at a large financial institution were subjects in this study, a pregnancy group and a control group. The first group consisted of 71 pregnant women. For each subject in the pregnancy group, a control group subject was randomly selected from the non-pregnant female employees with the same job title. The performance appraisal ratings were on a scale from 1 (outstanding performance) to 5 (unsatisfactory performance). The sample sizes, sample means and standard deviations for the two groups are given below:

Pregnancy group: $n_1 = 71$, $\bar{x}_1 = 2.38$, $s_1 = 1.10$

Control group: $n_2 = 71$, $\bar{x}_2 = 2.69$, $s_2 = 0.58$

The researchers hypothesized that pregnant employees would be rated differently when compared with the control group.

- Set up a null hypothesis and an alternative hypothesis to test whether the population mean performance ratings differed for the two groups of female employees.
- Calculate the t -test statistic and determine a p -value. State your conclusion.
- Calculate a 95% t -confidence interval for the difference in mean performance appraisal ratings for pregnant employees and non-pregnant female employees. On average, are the performance ratings for pregnant women better or worse than for the non-pregnant female employees? Explain.

2. Return to the study discussed in question 1. The same group of researchers also gathered data on the pregnant group's performance appraisal ratings during pregnancy and after returning from pregnancy leave. Here is a summary of the data gathered:

During pregnancy: $\bar{x}_1 = 2.38$, $s_1 = 1.10$, $n_1 = 71$

After pregnancy: $\bar{x}_2 = 2.65$, $s_2 = 1.64$, $n_2 = 71$

Difference “During – After”: $\bar{x}_d = -0.27$, $s_d = 2.00$, $n_d = 71$

a. The researchers were interested in answering the following question:

Did the mean performance ratings for the pregnancy group differ significantly between the During Pregnancy and After Pregnancy time periods?

Which test is more appropriate to answer this question, a one-sample t -test or a two-sample t -test? Explain.

b. Use the appropriate test to answer the question posed in (a). Report the value of the test statistic, the p -value, and your conclusion.

3. A state university is concerned that female students are not as well prepared in mathematics as their male counterparts. Random samples of 20 male students and 20 female students were selected from the class of first-year students. Their SAT Math scores are given below.

SAT Math Scores of Female Students

530 450 550 470 450 500 480 510 470 450
600 540 530 470 420 490 440 540 500 480

SAT Math Scores of Male Students

670 440 410 510 410 600 530 490 600 530
570 550 640 530 550 460 660 570 670 490

a. Make graphic displays to compare the SAT Math scores of the female students and the male students. Do your plots provide evidence that male students entering the university have higher SAT Math scores than female students?

b. Is it reasonable to assume that the distributions of SAT Math scores for both populations, first-year male students and first-year female students, are approximately normally distributed? Support your answer.

c. Calculate the sample means and standard deviations for the females and the males SAT Math scores.

d. Conduct a test of hypotheses to see if the mean SAT Math scores are significantly higher for males than for females. Report the value of the two-sample t -test statistic, the p -value and your conclusion.

4. A group of 4-year-olds, who were part of the Infant Growth Study, participated in a laboratory meal. Data collected during this meal can be used to answer the following research question: Do the mean number of calories consumed by girls at a meal differ from the mean number of calories consumed by boys? Below is a summary of the results:

Girls: $\bar{x}_G = 494$ calories, $s_G = 172$ calories, $n_G = 31$

Boys: $\bar{x}_B = 409$ calories, $s_B = 148$ calories, $n_B = 27$

a. Write the null and alternative hypotheses.

b. Calculate the value of the two-sample t -test statistic. (Round to three decimals.)

c. Adopt a conservative approach and set the degrees of freedom to be one less than the smaller of the two sample sizes. Calculate the p -value. Are the results significant at the 0.05 level? Explain.

d. For two-sample t -tests, the Content Overview of this unit suggested using a conservative approach for determining the degrees of freedom associated with the test statistic: set $df =$ smaller of $n_1 - 1$ or $n_2 - 1$, where n_1 and n_2 are the sample sizes of the two groups. However, statistical software calculates degrees for freedom using the following formula:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\left(\frac{1}{n_1 - 1} \right) \left(\frac{s_1^2}{n_1} \right)^2 + \left(\frac{1}{n_2 - 1} \right) \left(\frac{s_2^2}{n_2} \right)^2}$$

In general, this formula does not result in an integer value. In that case, the degrees of freedom are rounded down to the closest integer below the calculated value.

Use the formula given above to calculate the degrees of freedom. (Be sure to round down to the closest integer.) Calculate the p -value based on the df you calculated from the formula.

Based on this p -value, are the results significant at the 0.05 level?

REVIEW QUESTIONS

1. The financial aid office of a university asks a sample of students about their employment and earnings. The report says that “for academic year earnings, a significant difference ($p = 0.038$) was found between the sexes, with men earning more on the average. No difference ($p = 0.476$) was found between the earnings of African-American and white students.” Explain both of these conclusions, for the effects of sex and of race on mean earnings, in language understandable to someone who knows no statistics.

2. A study was conducted to investigate the effect of different levels of air pollution on the pulmonary functions of healthy, non-smoking, young men. Two geographical areas with different levels of air pollution were selected – Area 1 had lower levels of pollutants than Area 2. Samples of 60 men were selected from each area. The two groups of men had no significant differences in age, height, weight, and BMI. Data on two measures of pulmonary function for each group are provided below:

Forced vital capacity (FVC, in Liters)

$$\text{Area 1: } \bar{x}_1 = 4.49, s_1 = 0.43$$

$$\text{Area 2: } \bar{x}_2 = 4.32, s_2 = 0.45$$

Respiratory rate (RR, per minute)

$$\text{Area 1: } \bar{y}_1 = 17.17, s_1 = 4.26$$

$$\text{Area 2: } \bar{y}_2 = 16.28, s_2 = 2.39$$

- Why do you think the researchers tested to see if there were significant differences between the age, height, weight, and BMI for the two samples?
- Test whether there is a significant difference between the mean FVC for the participants from Area 1 and Area 2. State the value of the test statistic, the p -value and your conclusion.
- Test whether there is a significant difference between the mean RR for the two areas. State the value of the test statistic, the p -value and your conclusion.
- If you find a significant difference in (b) or (c) or both, construct a 95% confidence interval to estimate the difference in means between the two areas.

3. A state university is concerned that there is a difference in the writing abilities of their male and female students. To test this assertion, the university took a random sample of 60 of their first-year students and recorded their genders and SAT Writing scores. The data appears below.

SAT Writing Scores of Female Students

480 540 620 590 530 620 580 530 530 560 510 560 560
 550 520 480 560 510 500 540 490 430 610 620 510

SAT Writing Scores of Male Students

480 560 400 580 480 460 430 430 490 610 540 500 540
 400 530 640 350 470 600 610 530 580 430 510 520 380
 540 460 640 520 570 560 490 440 480

- a. Make comparative boxplots for the SAT Writing scores for female and male students. Based on your boxplots, is it reasonable to assume that SAT Writing scores are approximately normally distributed for each gender? Does one gender tend to have higher SAT Writing scores than the other?
- b. Summarize the data by reporting the sample sizes, sample means and standard deviations for both groups.
- c. Test to see if there is a significant difference in mean SAT Writing scores between female and male first-year students attending this university. Report the value of the test statistic, the p -value, and your conclusion.
- d. Compute a 95% confidence interval for the difference in mean SAT Writing scores for female and male students attending this university. Interpret your results.

4. Do 4-year-old boys eat, on average, more mouthfuls of food at a meal than 4-year-old girls? A group of 4-year-olds, who were part of the Infant Growth Study, participated in a laboratory meal.

Data on mouthfuls of food consumed during the laboratory meal were collected. Here is a summary of the results:

Girls: $\bar{x}_G = 80.8$ mouthfuls, $s_G = 41.4$ mouthfuls, $n_G = 31$

Boys: $\bar{x}_B = 92.3$ mouthfuls, $s_B = 42.0$ mouthfuls, $n_B = 27$

- a. State the null and alternative hypotheses.
- b. Determine the value of the two-sample t -test statistic and the p -value. Report your conclusion.