Unit 27: Comparing Two Means

PREREQUISITES

Students should have experience with one-sample *t*-procedures before they begin this unit. That material is covered in Unit 26, Small Sample Inference for One Mean.

ADDITIONAL TOPIC COVERAGE

Additional coverage of two-sample *t*-procedures can be found in *The Basic Practice of Statistics*, Chapter 19, Two-Sample Problems.

ACTIVITY DESCRIPTION

In this activity, students will check whether the mean number of chips per cookie in Nabisco's Chips Ahoy regular and reduced fat chocolate chip cookies differ. In Unit 25's activity, students collected data on the number of chips per cookie from Chips Ahoy regular chocolate chip cookies. They can use those data in this activity. In addition, if they have not already done so, they will need to collect data on the number of chips per cookie from one or more bags of Chips Ahoy reduced fat chocolate chip cookies.

MATERIALS

One or two bags each of Nabisco's Chips Ahoy chocolate chip cookies, regular and reduced fat; paper plates or paper towels.

If students have not collected the data as part of Unit 25's activity, then they should be assigned to work in groups of 2 to 4 students to collect the data for this activity. If you need to have a group of 3, students in the group can trade off being chip counters. See Unit 25's Activity Description for suggestions on collecting these data. The sample data, on which sample solutions to the activity are based, are given in Table T27.1. For the sample data, two

independent counts were taken on the number of chips in each cookie, and then the results were averaged.

Original Chip Count (averaged from two independent counts)

18.5	17.0	14.0	14.5	15.0	13.5	16.5	15.5	19.0	16.0	20.5
23.5	17.5	18.5	21.5	18.5	22.5	15.0	16.0	17.5	12.0	14.5
12.5	12.0	20.5	21.5	22.5	24.0	18.5	16.5	22.5	22.0	18.5
18.5	21.5	20.0	17.5	16.5	17.5	17.5	19.5	21.5	24.5	18.0
23.0	20.5	19.5	25.0	19.0	20.0	22.0	19.0	21.5	18.0	14.5
17.0	21.0	10.5	18.0	18.0	20.0	13.5	23.5	16.5	19.5	

Reduced Chip Count (averaged from two independent counts)

11.0	7.5	5.5	8.5	9.5	15.5	20.0	14.0	18.5	21.5	18.5
18.0	19.5	16.0	21.0	19.0	17.0	21.5	13.5	14.5	14.0	18.5
11.5	13.0	19.0	20.0	16.0	18.0	14.0	20.0	9.0	13.0	11.5
14.5	13.5	11.5	13.0	18.0	19.0	21.5	16.5	12.0	14.0	14.0
11.0	14.5	15.0	21.0	21.5	17.5	15.5	16.5	14.5	16.0	16.5
10.5	15.5	16.0	15.0	9.0	11.5	8.5	15.0	13.5	17.0	15.0
9.5	5.5	12.5	17.0	19.5	20.5	16.0	20.0	19.0	15.5	19.5

Table T27.1. Sample data.

Students will need a copy of the class data for questions 2 – 5. If you decide not to collect the data (a task students really enjoy), have the class use the sample data in Table T27.1, which was collected in two statistics classes (each class got a bag of each type of cookie).

THE VIDEO SOLUTIONS

1. Sample answer: Take a sample from all licensed drivers in some state and look at the number of tickets the males and females in this group received for moving violations. We could then compare the mean number of tickets for women to the mean number of tickets for men.

2. The Hadza are a group of traditional hunter-gatherers who live in a way that is very similar to our ancestors. The men hunt with bows and arrows and the women forage for berries and root vegetables.

3. The original assumption was that the Hadza would burn more calories than the Westerners do, due to their more active lifestyle.

4. A two-sample *t*-test was used.

5. It turned out that no significant difference was found between the mean daily total energy expenditures of the Hadza and the Westerners. So, the researchers' original assumption that the Hadza burned more calories than Westerners due to a more active lifestyle was not supported by the data.

6. He placed the blame on people eating too much rather than on a sedentary lifestyle.

UNIT ACTIVITY SOLUTIONS

1. See question 3.

2. a. Sample answer: Since chocolate chips have fat in them, we think that Nabisco may put fewer chips in its reduced fat cookies as a way of reducing the fat content.

b. Sample answer: Given our answer in (a), we decided to use a one-sided alternative. Let μ_1 = the mean number of chips per cookie in regular Chips Ahoy chocolate chip cookies and let μ_2 = the mean number of chips per cookie in reduced fat Chips Ahoy chocolate chip cookies. Below are the null and alternative hypotheses:

$H_0 : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 > 0$

3. Sample data:

Original Chip Count (averaged from two independent counts)

18.5	17.0	14.0	14.5	15.0	13.5	16.5	15.5	19.0	16.0	20.5
23.5	17.5	18.5	21.5	18.5	22.5	15.0	16.0	17.5	12.0	14.5
12.5	12.0	20.5	21.5	22.5	24.0	18.5	16.5	22.5	22.0	18.5
18.5	21.5	20.0	17.5	16.5	17.5	17.5	19.5	21.5	24.5	18.0
23.0	20.5	19.5	25.0	19.0	20.0	22.0	19.0	21.5	18.0	14.5
17.0	21.0	10.5	18.0	18.0	20.0	13.5	23.5	16.5	19.5	

Reduced Chip Count (averaged from two independent counts)

11.0	7.5	5.5	8.5	9.5	15.5	20.0	14.0	18.5	21.5	18.5
18.0	19.5	16.0	21.0	19.0	17.0	21.5	13.5	14.5	14.0	18.5
11.5	13.0	19.0	20.0	16.0	18.0	14.0	20.0	9.0	13.0	11.5
14.5	13.5	11.5	13.0	18.0	19.0	21.5	16.5	12.0	14.0	14.0
11.0	14.5	15.0	21.0	21.5	17.5	15.5	16.5	14.5	16.0	16.5
10.5	15.5	16.0	15.0	9.0	11.5	8.5	15.0	13.5	17.0	15.0
9.5	5.5	12.5	17.0	19.5	20.5	16.0	20.0	19.0	15.5	19.5

Regular: $n_1 = 65$, $\bar{x}_1 = 18.462$, $s_1 = 3.308$

Reduced Fat: $n_2 = 77$, $\bar{x}_2 = 15.266$, $s_2 = 3.941$

4. Based on comparative boxplots, it appears that Chips Ahoy regular has more chips per cookie than Chips Ahoy reduced fat.
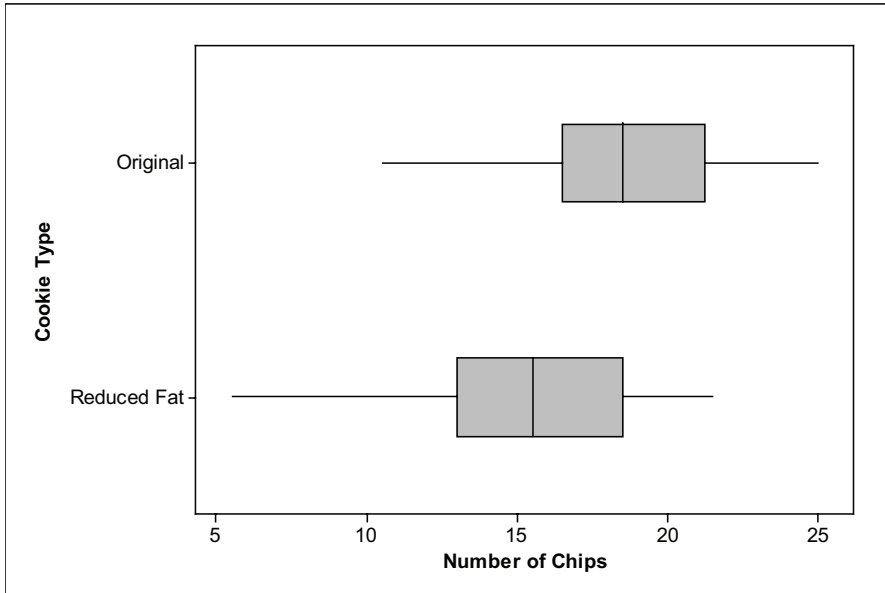
5. a. Sample answer:

$$t = \frac{(18.462 - 15.266) - 0}{\sqrt{\dfrac{3.308^2}{65} + \dfrac{3.941^2}{77}}} \approx 5.25$$

b. Sample answer: Using a conservative approach, we assume that $df = 65 - 1 = 64$. From software and using a one-sided alternative, we get a p-value that is essentially 0 (at least to 4 decimals).

c. Sample answer: Reject the null hypothesis. Conclude that the difference in the mean number of chips per cookie between the two types of cookies is positive. Hence, the mean number of chips per cookie in Chips Ahoy regular cookies is greater than the mean number of chips per cookie in Chips Ahoy reduced fat cookies.

6. Sample answer: $(18.462 - 15.266) \pm (1.998)\sqrt{\dfrac{(3.308)^2}{65} + \dfrac{(3.941)^2}{77}} \approx 3.196 \pm 1.215$, or

(1.981, 4.411)

EXERCISE SOLUTIONS

1. a. Let μ_1 be the mean job performance rating for pregnant employees and μ_2 be the mean job performance rating for non-pregnant female employees.

$H_0 : \mu_1 - \mu_2 = 0$

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

b. $t = \dfrac{(2.38 - 2.69) - 0}{\sqrt{\dfrac{1.10^2}{71} + \dfrac{0.58^2}{71}}} \approx -2.10$

Using a two-sided alternative, and letting df = 71-1 = 70 (the conservative approach), we determine p = 2(0.01967) ≈ 0.039. (If we use the statistical package Minitab, we get df = 106 and p = 0.038. The conservative approach yields a slightly higher p-value.) We conclude that there is a statistical difference in the mean job performance ratings for the two groups.

c. Using the conservative approach, we set df = 70 in order to determine the t-critical value: $t^* \approx 1.994$.

$$(2.38 - 2.69) \pm (1.994)\sqrt{\dfrac{1.10^2}{71} + \dfrac{0.58^2}{71}} \approx -0.31 \pm 0.294 \text{, or (-0.604, -0.016)}$$

Since both endpoints of the confidence interval are negative, the mean job performance appraisal rating for pregnant employees is lower than for non-pregnant female employees. In this case, lower ratings are associated with better job performance.

2. a. A one-sample paired t-test should be used. The "during" and "after" data involve the same group of women rather than samples from two independent populations.

b. $t = \dfrac{-0.27 - 0}{2.00 \big/ \sqrt{71}} \approx -1.138$;

based on a t-distribution with df = 70 and a two-sided alternative, p = 2(0.1295) ≈ 0.259. The mean difference in job performance appraisal ratings during pregnancy and after returning from pregnancy leave do not differ significantly.

3. a. Sample answer: It is somewhat difficult to tell if the distribution of SAT Math scores is higher for males than for females. From the dotplots, the distribution of SAT Math scores is more variable for males than for females. The highest SAT Math scores are from males but the lowest are also from males. Nevertheless, the midpoint of the distribution for the male students appears higher than the midpoint of the distribution for the female students.
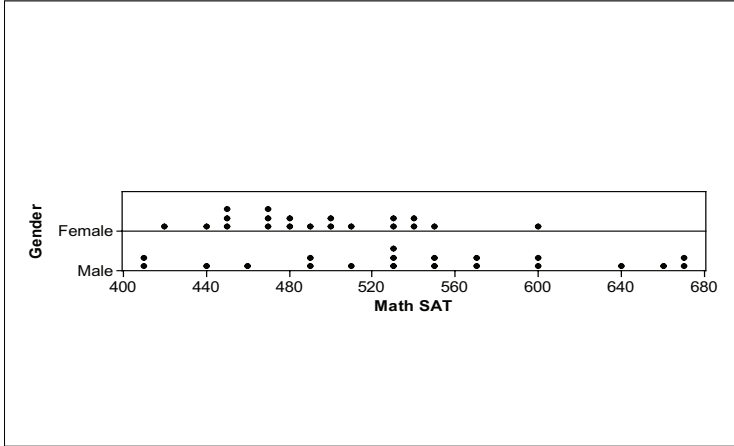
b. Consider the normal quantile plots below. Given the dots fall within 95% confidence interval bands on the normal quantile plots, it is reasonable to assume the data from both samples come from approximately normal distributions.
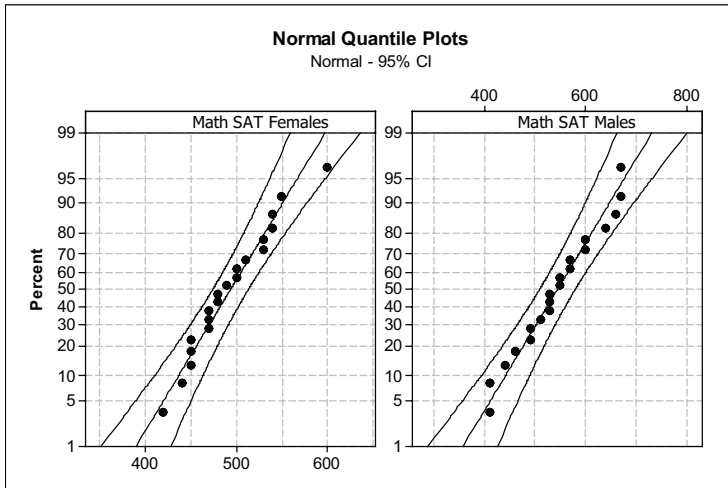
c. $\bar{x}_F = 493.5$, $s_F = 44.64$; $\bar{x}_M = 544.0$, $s_M = 80.6$

d. $t = \dfrac{(544.0 - 493.5) - 0}{\sqrt{\dfrac{80.6^2}{20} + \dfrac{44.64^2}{20}}} \approx 2.45$; using df = 19, we get the following p-value: $p \approx 0.012$

Therefore, we can conclude that the mean SAT Math scores are significantly higher for first-year male students entering this university than for first-year female students.

4. a. $H_0 : \mu_G - \mu_B = 0$ versus $H_a : \mu_G - \mu_B \neq 0$. (Could also be expressed as $H_0 : \mu_G = \mu_B$ versus $H_a : \mu_G \neq \mu_B$.)

b. $t = \dfrac{(494 - 409) - 0}{\sqrt{\dfrac{(172)^2}{31} + \dfrac{(148)^2}{27}}} \approx 2.023$

c. Use $df = 27 - 1$, or 26; $p = 2(0.02673) \approx 0.053$. Since $p > 0.5$, the results are not significant at the 0.05 level.

d. $df = \dfrac{\left(\dfrac{172^2}{31} + \dfrac{148^2}{27}\right)^2}{\left(\dfrac{1}{30}\right)\left(\dfrac{172^2}{31}\right)^2 + \left(\dfrac{1}{26}\right)\left(\dfrac{148^2}{27}\right)^2} \approx 55.9948$; use $df = 55$.

$p = 2(0.02397) \approx 0.048$. Since $p < 0.05$, the results are significant at the 0.05 level.

Notice that the p-value in (c) is only slightly larger using the conservative approach than it is here.

REVIEW QUESTIONS SOLUTIONS

1. The men in the sample earned more, on average, than did the women. The difference in the sample means was large enough that it couldn't be attributed to chance variation. If we assume that men and women in the entire student population have the same average earnings, the probability of observing a difference as large as we observed is only $p = 0.038$, making it pretty unlikely (less than a 4% chance). Because this probability is so small, we have good evidence that the average earnings of all male and female students (not just those who happen to be in the samples) differ.

There was also some difference between the average earnings of African- American and white students in the sample. However, a difference at least this large would happen almost half the time ($p = 0.476$) just by chance if African Americans and whites in the entire student population had exactly the same average earnings.

2. a. Sample answer: No mention was made that the samples were randomly selected from each area. The researchers wanted to be sure that the differences in pulmonary function could be attributed to the pollution levels in the two areas and not due to a lurking variable such as age, height, weight or BMI that was different for the men in the two groups of participants in the study.

b. $t = \dfrac{(4.49 - 4.32) - 0}{\sqrt{\dfrac{0.43^2}{60} + \dfrac{0.45^2}{60}}} \approx 2.116$; using $df = 59$, we get $p = 2(0.01929) \approx 0.039$

We conclude that there is a statistically significant difference in mean forced vital capacity (FVC) in healthy, non-smoking, young men in Area 1 and Area 2. (Because the sample size was large, even a relatively small difference in sample means turned out to be statistically different.)

c. $t = \dfrac{(17.17 - 16.28) - 0}{\sqrt{\dfrac{4.26^2}{60} + \dfrac{2.39^2}{60}}} \approx 1.411$; using $df = 59$, we get $p = 2(0.08175) \approx 0.164$

We conclude that there is no significant difference in the mean respiratory rate in healthy, non-smoking, young men in Areas 1 and 2.

d. We construct a 95% confidence interval for the difference in mean FVC between the two areas. We use $df = 59$ (conservative approach) to find the t-critical value: $t^* = 2.001$.

$$(4.49 - 4.32) \pm (2.001)\sqrt{\frac{0.43^2}{60} + \frac{0.45^2}{60}} \approx 0.17 \pm 0.16 \text{ or } (0.01, 0.33).$$

3. a. Both distributions appear reasonably symmetric. Neither data set has outliers. The whiskers on the boxplots appear slightly longer than half the width of the boxes. These are characteristics that you would expect from normal data. So, based on the boxplots it appears reasonable to assume that both data sets are approximately normal. Although the SAT Writing scores for the male students are more spread out than for the female students, it looks as if the SAT Writing scores for the female students tend to be higher than for the male students.
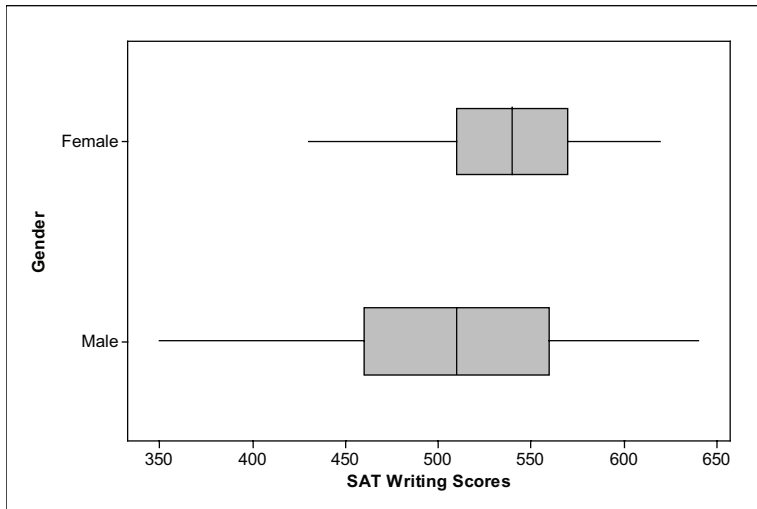
b. Female students: $n_F = 25$, $\bar{x}_F = 541.20$, $s_F = 48.59$

Male students: $n_M = 35$, $\bar{x}_F = 507.10$, $s_M = 73.70$

c. $t = \dfrac{(541.20 - 507.1) - 0}{\sqrt{\dfrac{(48.59)^2}{25} + \dfrac{(73.7)^2}{35}}} \approx 2.16$

Adopting a conservative approach, we use $df = 24$ to determine a p-value: We get $p = 2(0.02049) \approx 0.041 < 0.05$. We conclude that there is a significant difference between the mean SAT Writing scores for first-year women at this university compared to first-year men.

d. $(541.20 - 507.10) \pm 2.064\sqrt{\dfrac{48.59^2}{25} + \dfrac{73.7^2}{35}} \approx 34.10 \pm 32.61$, or (1.49, 66.71)

4. a. $H_0 : \mu_B - \mu_G = 0$ versus $H_a : \mu_B - \mu_G > 0$.

(Could also be expressed as $H_0 : \mu_B = \mu_G$ versus $H_0 : \mu_B = \mu_G$.)

b. $t = \dfrac{(92.3 - 80.8) - 0}{\sqrt{\dfrac{42.0^2}{27} + \dfrac{41.4^2}{31}}} \approx 1.05$; using $df = 26$, $p = 0.152$

There is insufficient evidence to reject the null hypothesis in favor of the alternative. Hence, we are unable to conclude from these data that 4-year-old boys consume more mouthfuls of food during a meal than 4-year-old girls.