Module 6 Comparing Two Populations (Means)

1. Do women tend to spend more time on housework than men? If so, how much more?

	Housework Hours		
Gender	**Sample Size**	**Mean**	**Standard Deviation**
Women	476	33.0	21.9
Men	496	19.9	14.6

(a) Based on this study, calculate how many more hours, on the average, women spend on housework than men.

(b) Find the standard error for comparing the means. What factors causes the standard error to be small compared to the sample standard deviations for the two groups?

(c) Calculate the 95% CI comparing the population means for women and men. Interpret the result including the relevance of 0 being within the interval or not.

Solution

(a) $33.0 - 19.9 = 13.1$

(b) $\text{se} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{21.9^2}{476} + \frac{14.6^2}{496}} = 1.20$. Because the sample size is large, the standard error is small compared to the sample standard deviations for the two groups.

(c) 95% CI is $(33.0 - 19.9) \pm 1.96 \times 1.2 = (10.75, 15.45)$. We can be 95% confident that the difference between the population mean housework hours of women and men falls between 10.75 and 15.45. Because 0 is not in the interval, we can conclude that there is a mean difference between the populations. It appears that the population mean for women is higher than the population mean for men.

2. A random sample of 100 students from MBA class made an average score of 60 with a standard deviation score of 15 in statistics. A random sample of 64 students from BS class made an average score of 66 with a standard deviation of 16 in the same course. Construct a 95% confidence interval for the difference between the mean score of the two classes.

Solution:
se $= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{15^2}{100} + \frac{16^2}{66}} = 2.48$
95% confidence interval: $(60 - 66) \pm 1.96 \times 2.48 = (-10.86, -1.14)$ where df=128. We can be 95% confident that the difference between the population mean score of MBA class and BS class falls between -10.86 and -1.14. Because 0 is not in the interval, we can conclude that there is a mean difference between the populations. It appears that the population mean score for MBA class is smaller than the population mean score for BS class.

3. Does Cell Phone Use While Driving Impair Reaction Times?
Experiment:

1. 64 college students
2. 32 were randomly assigned to the cell phone group
3. 32 to the control group students used a machine that simulated driving situations

At irregular periods a target flashed red or green Participants were instructed to press a brake button as soon as possible when they detected a red light. For each subject, the experiment analysed their mean response time over all the trials Averaged over all trials and subjects, the mean response time for the cell-phone group was 585.2 milliseconds. The mean response time for the control group was 533.7 milliseconds

Sample	N	Mean	StDev
Cell Phone	32	585.2	89.6
Control	32	533.7	65.3

Solution:

$$n_1 = 32, \bar{x}_1 = 585.2, s_1 = 89.6, n_2 = 32, \bar{x}_2 = 535.7, s_2 = 65.3$$

Step 1: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$.

Step 2: t-test: $t = \frac{585.2 - 533.7}{\sqrt{\frac{89.6^2}{32} + \frac{65.3^2}{32}}} = 2.63$.

Step 3: p-value: df=56. p-value=$P(t < -2.63) + P(t > 2.63) = 2P(t < -2.63) =$ 0.011

Step 4: The P-value is less than 0.05, so we can reject H_0. There is strong evidence to conclude that the population mean response times differ between the cell phone and control groups

4. Independent random samples of 17 sophomores and 13 juniors attending a large university yield the following data on grade point averages.

Sophomores			Juniors		
3.04	2.92	2.86	2.56	3.47	2.65
1.71	3.60	3.49	2.77	3.26	3.00
3.30	2.28	3.11	2.70	3.20	3.39
2.88	2.82	2.13	3.00	3.19	2.58
2.11	3.03	3.27	2.98		
2.60	3.13				

At the 5% significance level, do the data provide sufficient evidence to conclude that the mean GPAs of sophomores and juniors at the university differ?

Solution:

$$n_1 = 17, \bar{x}_1 = 2.84, s_1 = 0.52, n_2 = 13, \bar{x}_2 = 2.98, s_2 = 0.31$$

Step 1: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$.

Step 2: t-test: $t = \frac{2.84 - 2.98}{\sqrt{\frac{0.52^2}{17} + \frac{0.31^2}{13}}} = -0.92$.

Step 3: p-value: df=26. p-value=$P(t < -0.92) + P(t > 0.92) = 2P(t < -0.92) = 0.36$

Step 4: The P-value is larger than 0.05, so we cannot reject H_0. At 5% level of significance, the data does not provide sufficient evidence that the mean GPAs of sophomores and juniors at the university are different.

5. In a packing plant, a machine packs cartons with jars. It is supposed that a new machine will pack faster on the average than the machine currently used. To test that hypothesis, the times it takes each machine to pack ten cartons are recorded.

New machine					Old machine				
42.1	41.3	42.4	43.2	41.8	42.7	43.8	42.5	43.1	44.0
41.0	41.8	42.8	42.3	42.7	43.6	43.3	43.5	41.7	44.1
$\bar{y}_1 = 42.14, s_1 = 0.683$					$\bar{y}_2 = 43.23, s_2 = 0.750$				

Do the data provide sufficient evidence to conclude that, on the average, the new machine packs faster?

(a) Construct a 95% confidence interval comparing the population means for the new machine and the old machine.

(b) Perform a hypothesis test at the 5% level of significance.

Solution: Since s_1 and s_2 are not that different, the ratio of the two sample standard deviations is $\frac{s_1}{s_2} = \frac{0.683}{0.750} = 0.91$ which is quite close to 1. We can assume the population standard deviations are equal. Then the pooled sample standard deviation is:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(10 - 1)0.683^2 + (10 - 1)0.683^2}{10 + 10 - 2}} = 0.717$$

df=10+10-2=18

(a) The confidence interval is: $(42.14-43.23) \pm 2.10 \times 0.717\sqrt{\frac{1}{10} + \frac{1}{10}} = (-1.763, -0.417)$. We can be 95% confident that the difference between the population mean time of new machine and old machine take falls between -1.763 and -0.417. Because 0 is not in the interval, we can conclude that there is a mean difference between the populations. It appears that the new machine packs faster than old machine.

(b) Step 1: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 < 0$.

Step 2: t-test: $t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{42.14-43.23}{0.717\sqrt{\frac{1}{10}+\frac{1}{10}}} = -3.40$.

Step 3: p-value: df=18. p-value=$P(t < -3.40) = 0.002$

Step 4: The P-value is smaller than 0.05, so we can reject H_0. At 5% level of significance, the data provide sufficient evidence that the new machine packs faster than the old machine on average.

6. Hours spent studying per week were reported by students in a class survey. Students who say they usually sit in the front were compared to students who say they usually sit in the back.For the 99 students who reported that they usually sit in the front, the mean was 16.4 hours with a standard deviation of 10.85 hours. For the 94 students who reported that they usually sit in the back, the mean was 10.9 hours with a standard deviation of 8.41 hours. Both distributions were approximately normally distributed. Is the mean time spent studying for the two populations different?

(a) Construct a 95% confidence interval comparing the population mean time spent studying for the two populations.

(b) Perform a hypothesis test at the 5% level of significance.

Solution: Since s_1 and s_2 are not that different, the ratio of the two sample standard deviations is $\frac{s_1}{s_2} = \frac{10.85}{8.41} = 1.29 < 2$. We can assume the population standard deviations are equal. Then the pooled sample standard deviation is:

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(99-1)10.85^2 + (94-1)8.41^2}{99+94-2}} = 9.739$$

df=99+94-2=191

(a) The confidence interval is: $(16.4 - 10.9) \pm 1.972 \times 9.739\sqrt{\frac{1}{99}+\frac{1}{94}} = (4.93, 8.58)$. We can be 95% confident that the difference between the population mean time spent studying for the two populations falls between 4.93 and 8.58. Because 0 is not in the interval, we can conclude that there is a mean difference between the populations. It appears that the students sit in the front spent less studying time than the student sit in the back.

(b) Step 1: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$.

Step 2: t-test: $t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{16.4-10.9}{9.739\sqrt{\frac{1}{99}+\frac{1}{94}}} = 3.92$.

Step 3: p-value: df=191. p-value=$P(t < -3.92) + P(t > 3.92) = 2P(t < -3.92) = 0.0001$

Step 4: The P-value is smaller than 0.05, so we can reject H_0. At 5% level of significance, we decide that the mean time spent studying is different for the two populations.

7. Anna's project for her introductory statistics course was to compare the selling prices of textbooks at two Internet bookstores. She first took a random sample of 10 textbooks used that term in courses at her college, based on the list of texts compiled by the college bookstore. The prices of those textbooks at the two Internet sites were

$$\text{Site A: } \$115, \$79, \$43, \$140, \$99, \$30, \$80, \$99, \$119, \$69$$

$$\text{Site B: } \$110, \$79, \$40, \$129, \$99, \$30, \$69, \$99, \$109, \$66$$

(a) Are these independent samples or dependent samples? Justify your answer.

(b) Find the mean for each samples. Find the mean of differences scores. Compare, and interpret.

(c) Construct a significance test using a 0.1 significance level to compare the population mean prices of all textbooks used that term at her college.

Solution:

(a) Dependent samples. Price of same books on two different sites.

(b) $\bar{x}_A = \$87.30, \bar{x}_B = \$83.00$ and $\bar{d} = \$4.30$. The sample mean price for the books from site A is higher than the sample mean price for the books from site B. Thus the sample mean of the difference is positive.

(c) The differences are 5, 0, 3, 11, 0, 0, 11, 0, 10, 3. $n = 10, \bar{d} = 4.3, s_d = 4.715$

$$H_0 : \mu_d = 0 \text{vs} H_a : \mu_d \neq 0$$

$$t = \frac{4.3 - 0}{4.715/\sqrt{10}} = 2.88, df = 10 - 1 = 9$$

P-value= $P(t < -2.88) + P(t > 2.88) = 2P(t < -2.88) = 0.0182 < 0.05$, we reject H_0.