

Statistics 101 L – Homework 11

Solution

Assignment:

1. For each of the following situations, indicate whether it is a two-sample problem or a paired sample problem. Explain briefly your choice. Also indicate whether you would construct a confidence interval or perform a test of hypothesis.
 - a) A study is conducted to determine whether taking an ACT preparation course improves the overall score on the ACT. A random sample of 40 high school seniors from a large school district is selected. Half are assigned at random to go through a special six week ACT preparation course. The other half does nothing special for those six weeks. At the end of the six weeks, the students are given the ACT. The school district would like to determine if students completing the preparation course have significantly higher ACT scores, on average, than those who did nothing.

This is a two-sample problem because there are two separate groups of students (one group that does the ACT preparation course and the other that does nothing special). Perform a test of hypothesis comparing the mean scores on the ACT.

$$H_0 : \mu_{prep} = \mu_{no\ prep}$$

$$H_A : \mu_{prep} > \mu_{no\ prep}$$

- b) The effectiveness of a weight loss program that combines exercise and diet is to be evaluated. Thirty volunteers are weighed before beginning the program and re-weighed after following the program for three months. One wishes to estimate, with 95% confidence, the mean difference between the post-program and pre-program weights.

This is a paired sample problem because you have two measurements of weight on each individual in the program. Construct a 95% confidence interval.

- c) Over the years, the game show Jeopardy has had more male winners than female winners. One explanation for this phenomenon could be a difference in the reaction times between men and women (men buzz in faster than women). A random sample of 15 previous male contestants and a random sample of 15 previous female contestants are selected and a test of reaction times is given. The show's producers would like to determine if the mean reaction time of women is slower than the mean reaction time of men, on average.

This is a two-sample problem because the samples come from two separate populations (male contestants and female contestants). Perform a test of hypothesis comparing the mean times to buzz in.

$$H_0 : \mu_{men} = \mu_{women}$$

$$H_A : \mu_{men} < \mu_{women}$$

- d) In order to compare computer speeds, a set of benchmark programs are run and the CPU time required to complete each program is recorded. Six benchmark programs are selected. Each program is run on two computers and the CPU times are recorded. One wishes to estimate, with 98% confidence, the mean difference in CPU times for the two computers.

This is a paired sample because there are two CPU times for each program (the times are paired by program). Construct a 98% confidence interval.

2. A major medical center in the Northeastern U.S. conducted a study looking at blood cholesterol levels and incidence of heart attack. Below are data from 16 people who had a heart attack and 20 people who did not have a heart attack.

(1) Heart Attack				(2) No Heart Attack				
242	186	266	206	182	222	198	192	238
318	294	282	234	198	188	166	204	182
224	276	262	360	178	202	164	230	186
310	280	248	258	162	182	218	170	200

- a) Is this an experiment or an observational study? Explain briefly.

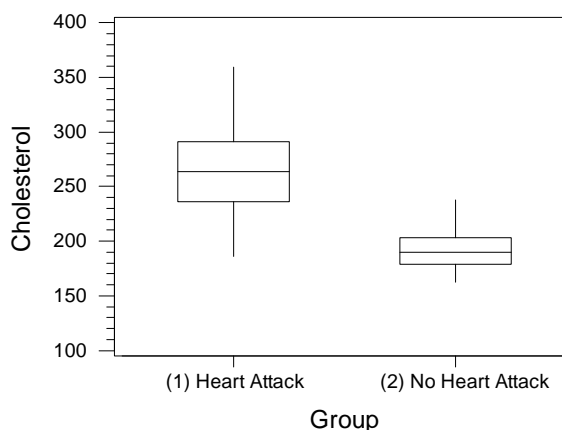
This is an observational study. No factor, or variable, is being manipulated. The cholesterol levels are being observed for the two naturally occurring groups.

- b) Compute 5-number summaries for each group.

Group	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
(1) Heart Attack	186	238	264	288	360
(2) No Heart Attack	162	180	190	203	238

- c) Construct side-by-side box plots. Compare the two groups in terms of center and spread.

Side-by-side box plots indicate that the heart attack group members have a higher central level of cholesterol and more spread in their cholesterol values than the no heart attack group.



- d) Describe how the individuals in the study need to be selected in order for the randomization condition to be met.

The 16 individuals should be selected at random from a population of individuals who have experienced heart attacks. The 20 individuals should be selected at random from a population of individuals who have not experienced a heart attack.

Below are summary statistics for the two groups.

(1) Heart Attack	(2) No Heart Attack
$\bar{y}_1 = 265.375$	$\bar{y}_2 = 193.1$
$s_1 = 43.645$	$s_2 = 21.623$
$n_1 = 16$	$n_2 = 20$

- e) Is there sufficient evidence to indicate that the mean cholesterol for people who have had a heart attack is greater than that for people who have not had a heart attack? Perform the appropriate test of hypothesis. Use $df = 20$ and $\alpha=0.05$. Note: the nearly normal condition is satisfied.

$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 > \mu_2$$

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(265.375 - 193.1)}{\sqrt{\frac{(43.645)^2}{16} + \frac{(21.623)^2}{20}}} = \frac{72.275}{11.9345} = 6.056$$

Using Table T and 20 degrees of freedom, the one tail probability (P-value) is less than 0.005. Because the P-value is so small, we reject the null hypothesis. People who have had a heart attack have a population mean cholesterol level that is higher than that of people who have not had a heart attack.

- f) Construct a 95% confidence interval for the difference in population mean cholesterol levels.

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

$$t^* = 2.086$$

$$72.275 \pm 2.086(11.9345)$$

$$72.275 \pm 24.895$$

$$47.38 \text{ to } 97.17$$

We are 95% confident that the population mean cholesterol for people who have had a heart attack is from 47.38 to 97.17 points higher than that for people who have not had a heart attack.

3. As part of a study on pulse rates some participants (randomly assigned as controls) had their pulse rate taken and asked to sit quietly for 5 minutes. At the end of 5 minutes, their pulse rates were taken again. Below are the data.

Participant	1	2	3	4	5	6	7	8	9	10
Pulse rate 1	86	71	90	68	71	68	74	70	78	69
Pulse rate 2	88	73	88	72	77	68	76	71	82	67

- a) Why is this a paired sample problem?

This is a paired sample problem because there are two measurements (pulse rates) for each participant. There is a pair of pulse rates for each participant.

- b) Based on the JMP output below, is the population mean difference zero or is it something different? You must take the information from the JMP output and use it to provide supporting evidence for your answer to this question.

Look at the difference (Pulse rate 2 – Pulse rate 1) in pulse rates.

The conditions appear to be met. The participants were randomly assigned to the group so the randomization condition is met. This appears to be part of an experiment so the 10% condition does not apply. The paired sample assumption is discussed in a) above. For the nearly normal condition, the histogram shows a fairly symmetric, mounded in the middle distribution. The box plot is symmetric. The points on the normal quantile plot fall close to the red diagonal line representing the normal model. These differences could have come from a population described by a normal model.

$$H_0 : \mu_d = 0$$

$$H_A : \mu_d \neq 0$$

$$t = \frac{\bar{d} - 0}{\left(\frac{s_d}{\sqrt{n}} \right)} = 2.0803 \quad P - \text{value} = 0.0672$$

Because the P-value is not small (it is greater than 0.05), we fail to reject the null hypothesis. The population mean difference in pulse rates could be zero.

Alternatively, the 95% confidence interval for the population mean difference goes from –0.15 to 3.55. Because zero is in this confidence interval, zero is a plausible value for the population mean difference. That is the population mean difference in pulse rates could be zero.