

STAT 2507

Assignment # 5 Solutions

Due: Thursday, December 2, 2014, in class

Last Name \_\_\_\_\_ First \_\_\_\_\_

Student Number \_\_\_\_\_ Lab Section (Important) \_\_\_\_\_

Total of marks=100. Marks for each question are given in [ ]

**Part I. Minitab questions. Use only the blanks left to answer lab questions.**

1. Data in column A (of the Excel file) corresponds to amounts in dollar a sample of 30 kids gave to a local charity during its last campaign. Assuming data coming from a normal population with standard deviation 3, use minitab to run a z test at significance level  $\alpha = .05$  to check if the average donation is different from 6. You can proceed as follows: copy (from Excel file that comes with the assignment 5) and paste column A into column C1 (Have a minitab session opened), then type the following commands:

onez c1;  
sigma 3;  
test 6.

- (a) [3 marks] What is the value of your test statistic? \_\_\_\_\_

(Answer:  $Z = 0.55$ .)

- (b) [3 marks] What is the P-value of your test? \_\_\_\_\_

(Answer:  $p - value = 0.584$ .)

- (c) [6 marks] What is your conclusion regarding the average donation? \_\_\_\_\_

\_\_\_\_\_  
(Answer: The data do not present sufficient evidence to indicate that the average donation is different from 6 dollars.)

2. Data in column B (of the Excel file) corresponds to waiting times in minutes of a sample of 25 customers when they called a hotline service. Use minitab to run a t test to see if the average waiting time exceeds 7 minutes as one of customers complains. You can proceed as follows: Copy and paste column B (Excel) into column C2 (minitab), then type the following commands:

onet c2;  
test 7;  
alternative +1.

- (a) [4 marks] What is the value of your test statistic? \_\_\_\_\_

(Answer:  $T = 3.82$ .)

- (b) [4 marks] What is the P-value of your test? \_\_\_\_\_

(Answer:  $p - value = 0.000$ .)

- (c) [5 marks] What conclusion about the customer's claim do you make? \_\_\_\_\_

---

(Answer: The data present sufficient evidence to support the customers claim that the mean waiting time exceeds 7 minutes.)

## Part II. Long-answer questions

1. A fast food franchiser is considering building a restaurant at a certain location. According to a financial analysis, a site is acceptable only if the number of pedestrians passing the location averages more than 100 per hour. A random sample of 25 hours produced  $\bar{x} = 108$  and  $s = 30$  pedestrians per hour.

- (a) [10] Do these data provide sufficient evidence to establish that the site is acceptable? Use  $\alpha = 0.05$ .

**Solution:** Let  $\mu$  be the true average of the pedestrians who pass the location in one hour. The Hypotheses are  $H_0 : \mu = 100$  vs  $H_a : \mu > 100$ . Test statistic  $z = \frac{108-100}{30/\sqrt{25}} \approx 1.333$ . The critical point for this problem is  $z_{.05} = 1.645$ . Since, 1.333 is not greater than 1.645 we do not reject  $H_0$  in vs favor of  $H_a$ . In other words, there is no evidence at level  $\alpha = 0.05$  that the true average of number pedestrians per hours is more that 100.

**or**

$t = \frac{108-100}{30/\sqrt{25}} \approx 1.333$   $df = 24$  and right-tailed test  $\Rightarrow t_{0.05} = 1.711$ . Since 1.333 is not greater than 1.711, we do not reject  $H_0$ . In other words, there is no evidence at level  $\alpha = 0.05$  that the true average of number pedestrians per hours is more that 100.

- (b) [6] What are the consequences of Type I and Type II errors? Which error is more expensive to make?

**Solution:** Type I error is rejecting  $H_0$  when it is true and the result of this error is to construct the site when there are not enough pedestrians passing the location. (bankruptcy). Type II error is to accept  $H_0$  when it is not true, so the site does not construct in location more than 100 pedestrians. (missing a good location). Type I error is more expensive. (It is possible type II error be more expensive. Accept if it supports with good explanation).

- (c) [4] Considering your answer in part (b), if you could select an  $\alpha$ , would you select it to be large or small? Explain.

**Solution:**  $\alpha$  should be small. The smaller  $\alpha$  is the less likely it is for Type I error, i.e., rejecting  $H_0$  falsely, to happen.

2. Quebec was believed to be different from other provinces regarding the rate of deaths due to cancer. In order to test whether this belief is true, two independent samples were taken and summarized as follows:

Quebec:  $n_1 = 150$ , Number of deaths due to cancer:  $x_1 = 47$ .

Rest of Canada:  $n_2 = 1000$ , Number of deaths due to cancer:  $x_2 = 291$ .

Let  $p_1$  and  $p_2$  denote the rates of deaths due to cancer for Quebec and the rest of Canada, respectively. The significance level for this test is  $\alpha = 0.05$ .

- (a) [2] Using  $p_1$  and  $p_2$ , write the null and alternative hypotheses.

**Solution:** Null hypothesis:  $H_0 : p_1 - p_2 = 0$ , Alternative hypothesis:  $H_a : p_1 - p_2 \neq 0$ .

(b) [8] Test this using p-value approach.

**Solution:**  $\hat{p}_1 = 147/150 = 0.3133$ ,  $\hat{p}_2 = 291/1000 = 0.2910$ ,  $\hat{p} = (47 + 291)/(150 + 1000) = 0.294$ . Now,  $Z = (\hat{p}_1 - \hat{p}_2)/\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)} \doteq 0.56$ , Thus, since this is a two-tailed test,  $p - value \doteq P(Z < -0.56) + P(Z > 0.56) = 0.2877 + (1 - 0.7123) = 0.5754$ . Since  $p - value > \alpha = 0.05$ ,  $H_0$  is not rejected.

(c) [8] Test this using the critical value approach.

**Solution:** From the standard normal table,  $z_{\alpha/2} = z_{0.025} = 1.96$ . Since  $-1.96 < Z (= 0.56) < 1.96$ , we do not reject  $H_0$ .

(d) [2] Do the conclusions you made in (b) and (c) agree?

**Solution:** Yes, they do agree.

3. [15] A company is interested in offering its employees one of two employee benefit packages. A random sample of the company's employees is collected, and each person in the sample is asked to rate each of the two packages on an overall preference scale of 0 to 100. Results were

Employee	Program A	Program B
1	45	40
2	67	70
3	63	60
4	59	45
5	77	85
6	69	79
7	45	50
8	39	46
9	52	50
10	58	60
11	75	80

Do you believe that, at level  $\alpha = 0.05$ , the employees of this company prefer, in average, one package over the other? Explain your answer by stating an appropriate two sided hypothesis testing problem.

**Solution:** We wish to test  $H_0 : \mu_1 - \mu_2 = 0$  vs.  $H_a : \mu_1 - \mu_2 \neq 0$ , or using paired different test,  $H_0 : \mu_d = 0$  vs.  $H_a : \mu_d \neq 0$

It is paired data where  $d_i : 5, -3, 3, 14, -8, -10, -5, -7, 2, -2, -5$

$$\bar{d} = \sum d_i/n = -39/11 = -1.4545 \text{ and } s_d^2 : \frac{\sum d_i^2 - (\sum d_i)^2/n}{n-1} = \frac{725 - 138.27}{10} = 48.672$$

The test statistic is  $t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{-1.4545 - 0}{6.976/\sqrt{11}} = -0.6914$  and the critical value is  $-t_{0.025;10} = -2.228$ , so  $H_0$  is not rejected. In other words, at level 0.05, the employees do not prefer one package over the other.

4. [20] In an attempt to compare the starting salaries for university graduates who majored in education and the social sciences, random samples of 100 recent university graduates were selected from each major and the following sample information was obtained:

Major	Mean	St. Dev.
Education	\$52,554	\$3225
Social Science	\$49,348	\$2475

Conduct an appropriate hypothesis test at the 5% level of significance to determine if there is a difference in the average starting salaries for all university graduates who majored in education and the social sciences. Conduct this test using

(a) [7] the p-value method

**Solution:** Let  $\mu_1$  be the population mean for salary of graduates who majored in Education;  $\mu_2$  be the population mean for salary of graduates who majored in Social Sciences.

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

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

Distribution of salary of neither the Education graduates nor Social Science graduates is given. But,  $n_1 = 100 > 30$  and  $n_2 = 100 > 30$ . Hence, the large sample statistical test for  $(\mu_1 - \mu_2)$  gives the following test statistic.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{52554 - 49348}{\sqrt{\frac{3225^2}{100} + \frac{2475^2}{100}}} = \frac{3206}{406.525} = 7.89$$

. This is a two-tailed test. Hence,

$$p - value = P(Z > 7.89) + P(Z < -7.89) \approx 0 + 0 = 0$$

$\Rightarrow p - value = 0 < \alpha = 0.02$ . Hence, null hypothesis,  $H_0$ , is rejected.

(b) [7] the critical value method

**Solution:** This is a two-tailed test with significance level,  $\alpha = 0.05$ . Hence,  $Z_{0.025} = 1.96$   $\Rightarrow$  critical values are -1.96 and 1.96. Since the value of test statistic  $Z = 7.89 > 1.96$  the null hypothesis is rejected.

(c) [6] the confidence interval method

**Solution:** Sample sizes are greater than 30. Hence, 95% large sample confidence interval for difference in population mean  $(\mu_1 - \mu_2)$

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} &= 52554 - 50348 \pm Z_{0.025} \sqrt{\frac{3225^2}{100} + \frac{2475^2}{100}} \\ &= 3206 \pm (1.96)(406.525) = 3206 \pm 796.789 = 3206 \pm 796.79 \\ &2409.21.797 < (\mu_1 - \mu_2) < 4002.789 \end{aligned}$$

Since the confidence interval does not enclose  $(\mu_1 - \mu_2) = 0$ , null hypothesis,  $H_0$ , is rejected.