STAT 2507 Assignment # 5 Fall 2013

Last Name ——————————————- , **First** ——————-

Student # ——————
Lab Session ——————

<u>**Due**</u> **December 9, IN CLASS,**

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

Part I. Lab 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 20 kids gave to a local charity during its last campaign. Assuming data coming from a normal population with standard deviation 4, use minitab to run a z test at significance level $\alpha = .05$ to check if the average donation is <u>different from 5</u>. 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 4;
   test 5.

   (a) [5 marks] What is the P-value of your test?**.199**
   (b) [7 marks] What is your conclusion regarding the average donation?**We keep H_0. No evidence that average donation is different from $6**

2. Data in column B (of the Excel file) corresponds to waiting times in minutes of a sample of 100 customers when they called a hotline service. Use minitab to run a t test to see if the average waiting time is <u>below 6</u> minutes as the service provider claims. You can proceed as follows: Copy and paste column B (Excel) into column C2 (minitab), then type the following commands:
   onet c2;
   test 6;
   alternative -1.

   (a) [5 marks] What is the P-value of your test?**.016**
   (b) [8 marks] What conclusion about the provider's claim do you make?**We reject H_0. There is strong evidence that the waiting time is below 6 min, as the provider claims.**

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 50 hours produced $\bar{x} = 110$ and $s = 12$ pedestrians per hour.

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

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

(c) [4] Considering your answer in part (b), should you select α to be large or small? Explain.

Solution: **a.** $H_0 : \mu = 100$ vs. $H_a : \mu > 100$, since n is large enough, test statistic is $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{110 - 100}{12/\sqrt{50}} = 5.89$, null hypothesis is rejected since $5.89 > 1.645 = z_{0.05}$
b. 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. α should be a small number.

2. An experiment was conducted to test the effect of a new drug on a viral infection. The infection was induced in 100 mice, and the mice were randomly split into two groups of 50. The first group, the *control group*, received no treatment for the infection. The second group received the drug. After a 30-day period, the proportions of survivors, $\hat{p}_1$ and $\hat{p}_2$, in the two groups were found to be 0.36 and 0.60, respectively. Is there sufficient evidence to indicate that the drug is effective in treating the viral infection? State the null and the alternative hypothesis and test at 5% significance level using
(a) [9] P-value approach
(b) [9] Critical value approach
(c) [2] Do the conclusions you made in (a) and (b) agree?

Solution:
$H_0 : p_1 - p_2 = 0$ versus $H_a : p_1 - p_2 < 0$, where p_1 and p_2 are, respectively, the actual survivor proportions of the control and the treatment groups.
(a) First note that the number of survivors is the two samples are $x_1 = 0.36 * 50 = 18$ and $x_2 = 0.6 * 50 = 30$. Therefore, $\hat{p} = (18 + 30)/100 = 0.48$
The test statistic is $z = \frac{0.36 - 0.6 - 0}{\sqrt{(0.48)(0.52) * \frac{2}{50}}} = -2.401$
$P - value = P(Z < -2.401) \approx 0.0082 \leq 0.05$ we reject H_0
(b) The critical value is $-z_{.05} = -1.65$. Since $-2.401 < -1.65$ we reject H_0.
(c) The two conclusions should always agree as the critical value and p-value approach always lead to same conclusion.

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	56
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	70	82

Do you believe that the employees of this company prefer, on the average, one package over the other? Use $\alpha = 0.05$ 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 $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$

It is paired data where $d_i : -11, -3, 3, 14, -8, -10, -5, -7, 2, -2, -12$

$\bar{d} = \sum d_i/n = -39/11 = -3.55$ and $s_d^2 : \frac{\sum d_i^2 - (\sum d_i)^2/n}{n-1} = \frac{725-138.27}{10} = 58.67$

The test statistic is $t = \dfrac{\bar{d} - 0}{s_d/\sqrt{n}} = \dfrac{-3.55 - 0}{7.66/\sqrt{11}} = -1.54$ and $-t_{0.025}(10) = -2.228$, so H_0 is not rejected.

4. 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	\$50,554	\$2225
Social Science	\$48,348	\$2375

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 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

(b) [7] the critical value method

(c) [6] the confidence interval method .

Solution: Let μ_1 be the average starting sealery of Education graduates

$\mu_2 =$ the average starting sealery of Education graduates. We wish to test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$. Observe that in this problem the sample sizes $n_1, n_2 \geq 30$. Hence the test statistics is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{50554 - 48348}{\sqrt{\frac{(2225)^2}{100} + \frac{(2375)^2}{100}}} = 6.8$$

(a) $p - value = P(Z > 6.8) + P(Z < -6.8) \approx 0 + 0 = 0$ (we use the normal table here!). Since $p - value < 0.05 = \alpha \rightarrow$ we reject H_0.
(b) The critical values here are $z_{.025} = 1.96$ and $-z_{.025} = -1.96$ (from the normal table). Since the value of the test statistic is 6.8 and $6.8 > 1.96$, we reject H_0.
(c) The 95% confidence interval is

$$\left(\bar{x}_1 - \bar{x}_2 \pm 1.96\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = \left(2206 \pm 1.96 \times 324.44\right) = \left(1568.137, 2843.862\right).$$

Since, 0 is not in the confidence interval $\left(1568.137, 2843.862\right)$, we reject H_0 in favor of H_a.