STAT 2507 Assignment #5, Fall 2012

Section A due Tuesday, November 27, 2012 before 10:05am

Section B due Tuesday, November 27, 2012 before 6:05pm

Last Name __________ ***First Name*** __________ ***Student Number*** __________

Instructions

- *Use only the blanks left to answer lab questions. DO NOT print the data you are asked to generate.*
- *Total Marks=50.*

Part I Lab Questions

Question 1. [*Hypothesis testing for μ when σ is known*] Imagine choosing $n = 16$ woman at random from a large population and measuring their heights. Assume that the heights of the women in this population are normal with $\mu = 63.8$ inches and $\sigma = 3$ inches. Suppose you want to test the null hypothesis

$$H_0: \ \mu = 63.8 \ \text{ versus } \ H_a: \ \mu \neq 63.8$$

using $\alpha = 0.1$. Simulate the results of doing this test 30 times as follow:

random 16 c1-c30;

normal 63.8 3.

ztest 63.8 3 c1-c30

(a) [1] In how many tests did you reject H_0? That is, how many times did you make an "incorrect decision"? ____________________.

(b) [1] Are the p-values all the same for the 30 tests? __________.

(c) [2] Suppose you used $\alpha = 0.01$ instead of $\alpha = 0.1$. Does this change any of your decisions to reject or not? __________. In general, should the number of rejections increase or decrease if $\alpha = 0.01$ is used instead of $\alpha = 0.1$? __________.

Solution:

(a) Any number less than 30.
(b) No.
(c) Yes, Decrease.

 Continued on next page...

Question 2. [*Hypothesis testing for μ when σ is unknown*] Imagine choosing $n = 16$ woman at random from a large population and measuring their heights. Assume that the heights of the women in this population are normal with $\mu = 63.8$ inches and $\sigma = 3$ inches. Suppose you want to test the null hypothesis

$$H_0: \mu = 63.8 \text{ versus } H_a: \mu \neq 63.8$$

using $\alpha = 0.1$. Simulate the results of doing this test 30 times as follow:

random 16 c1-c30;
normal 63.8 3.
ttest 63.8 c1-c30

(a) [1] In how many tests did you reject H_0. That is, how many times did you make an "incorrect decision"? ___________________.

(b) [1] Are the p-values all the same for the 30 tests? __________.

(c) [2]Suppose you used $\alpha = 0.01$ instead of $\alpha = 0.1$. Does this change any of your decisions to reject or not? __________. In general, should the number of rejections increase or decrease if $\alpha = 0.01$ is used instead of $\alpha = 0.1$.__________.

Solution:

(a) Any number less than 30.

(b) No.

(c) Yes, Decrease.

 Continued on next page...

Part II Written Questions

Question 3. [3] A parent believes the average height for 14 year old girls differs from that of 14 year old boys. Estimate the difference in the height between girls and boys using a 95% confidence interval. The summary data are listed below where height is in feet.

$$n_1 = 40,\ \overline{x}_1 = 5.1,\ s_1 = 0.2,\ n_2 = 40,\ \overline{x}_2 = 4.8,\ s_2 = 0.3.$$

Based on your interval, do you think there is a significant difference between the true mean height of 14 year old girls and boys? Explain.

Solution:

$$(\overline{x}_1 - \overline{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (5.1 - 4.8) \pm (1.96)(.057) = (0.19,\ 0.41).$$

or

$$(\overline{x}_2 - \overline{x}_1) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (4.8 - 5.1) \pm (1.96)(.057) = (-0.41,\ -0.19).$$

Since this interval does not contain zero, the sample data support the claim that the heights of girls and boys differ.

Question 4. [4] In a study of the relationship between birth order and college success, an investigator found that 140 in a sample of 200 college graduates were firstborn or only children. In a sample of 120 non-graduates of comparable age and socioeconomic background, the number of firstborn or only children was 66. Estimate the difference between the proportions of firstborn or only children in the two populations from which these samples were drawn. Use a 90% confidence interval and interpret your results.

Solution: Denote by p_1 and p_2 the proportions for the population of college graduates and the population of non-graduates, respectively. We have

$$\widehat{p}_1 = 140/200 = 0.70, \text{ and } \widehat{p}_2 = 66/120 = 0.55.$$

Therefore the approximate 90% confidence interval is

$$\begin{aligned} &(\widehat{p}_1 - \widehat{p}_2) \pm 1.645\sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_2}} \\ =&(0.70 - 0.55) \pm (1.645)(0.0558) \\ =&0.15 \pm 0.092 \\ \text{or }& (0.058, 0.242). \end{aligned}$$

One can estimate with 90% confidence that the difference between the proportions of firstborn or only children in the two populations from which these samples were drawn is roughly between 0.06 and 0.24. In repeated sampling, all intervals constructed in this manner will enclose $(p_1 - p_2)$ 95% of the time. Hence, we are fairly certain that this particular interval encloses . Moreover, since the value $p_1 - p_2 = 0$ is not in the confidence interval, it is not likely that $p_1 = p_2$. We should conclude that there is a difference in the proportions of firstborn or only children in the two populations.

 Continued on next page...

Question 5. The length of duration, in minutes, of earthquakes in British Columbia has been recorded for future analysis and information. The length of duration of a random sample of 6 earthquakes are as follows:

$$1.1, 0.9, 1.5, 0.7, 1.4, 1.3.$$

(a) [2] Assuming the distribution of the length of duration of the earthquakes is approximately normal, find a 98% confidence interval for the true average duration of earthquakes in British Columbia.

(b) [1] Interpret the interval in the previous question.

(c) [1] An earthquake expert claims that the average duration of earthquakes in British Columbia is 0.5 minutes. Based on the interval in part (a), can this claim be rejected? Justify your answer.

Solution:

(a) The sample mean and standard deviation are

$$\mu = 1.15, \ s = 0.308.$$

The 98% confidence interval is:

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 1.15 \pm 3.365\frac{0.308}{\sqrt{6}} = 1.15 \pm 0.423, \quad \text{or} \quad 0.727 < \mu < 1.573.$$

(b) One can estimate with 98% confidence that the true average duration of earthquakes in British Columbia is between 0.727 and 1.573 minutes.

(c) Since 0.5 is outside the limits of the 98% confidence interval, one can reject the experts claim that the true average duration of earthquakes in California is 0.5 minutes.

Question 6. A machine shop is interested in determining a measure of the current years sales revenue in order to compare it with known results from last year. From the 9682 sales invoices for the current year to date, the management randomly selected invoices and from each recorded x, the sales revenue per invoice. Using the following data summary, test the hypothesis that the mean revenue per invoice is \$6.35, the same as last year, with $\alpha = 0.05$.

$n = 400$	$\sum_{i=1}^{400} x_i = \$2464.40$	$\sum_{i=1}^{400} x_i^2 = 16,156.728$

(a) [1] State the null and alternative hypothesis to be tested.

(b) [2] Conduct a statistical test of the null hypothesis and give your conclusion.

(c) [1] Calculate the p-value for the test statistic in part (b).

Solution: Denote by μ the current mean revenue per invoice.

(a) $H_0: \ \mu = 6.35$, vs. $H_a: \ \mu \neq 6.35$

 Continued on next page. . .

(b) The observed value of the test statistic is:

$$\frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{6.161 - 6.35}{1.56/20} = -2.42$$

where

$$\overline{x} = 2460.40/400 = 6.161$$

$$s^2 = \frac{\sum_{i=1}^{400} x_i^2 - \frac{\sum_{i=1}^{400} x_i}{n}}{n-1} = \frac{16156.728 - 15183.168}{399} = 2.44$$

$$s = \sqrt{s^2} = \sqrt{2.44} = 1.56$$

Since $|z| = 2.42 > 1.96 = z_{0.025}$, we would reject H_0 at the 5% significance level. Mean revenue is different from 6.35.

(c) p-value$=P(|Z| > 2.42) = 0.0156$.

Question 7. To investigate a possible built-in sex bias in a graduate school entrance examination, 50 male and 50 female graduate students who were rated as above-average graduate students by their professors were selected to participate in the study by actually taking this test. Their test results on this examination are summarized in the following table. Do these data indicate that males will, on the average, score higher than females of the same ability on this exam? Use $\alpha = 0.01$.

	Males	Females
$\overline{x}$	720	693
s^2	104	85
n	50	50

(a) [1] State the null and alternative hypothesis to be tested.
(b) [2] Conduct a statistical test of the null hypothesis and state your conclusion.
(c) [1] Calculate the p-value for the test statistic in part (b).

Solution: Denote by μ_1 the average score for male students, and by μ_2 the average score for female students.

(a) $H_0: \ \mu_1 - \mu_2 = 0$, vs. $H_a: \ \mu_1 - \mu_2 > 0$;
(b) The observed value of the test statistic is:

$$\frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{720 - 693}{\sqrt{\frac{104}{50} + \frac{85}{50}}} = 13.92 > 2.33$$

Thus, we reject H_0 with $\alpha = 0.01$ and conclude that there is a significant difference in the mean scores. Men score, on the average, higher than women.

(c) p-value$=P(|Z| > 13.92) \approx 0$.

Question 8. A gas company president for a particular city is interested in the proportion of homes heated by gas. Historically, the proportion of homes heated by gas has been 0.65. A sample of 75 home was selected and it was found that 44 of them are heated with gas.

 Continued on next page...

(a) [4] Perform the appropriate test of hypothesis to determine whether the proportion of home heated by gas has changed. Test using $\alpha = 0.1$.
(b) [1] Find the p-value for this test.

Solution:

(a) The sample proportion of homes heated by gas is $\widehat{p} = x/n = 44/75 = 0.587$. Let p = the true proportion of homes heated by gas. The hypotheses to be tested are:

$$H_0: \ p = 0.65 \text{ vs. } H_a: p \neq 0.65$$

The value of test statistics is:

$$z = \frac{0.587 - 0.65}{0.0551} = -1.14.$$

Since $-1.645 < z < 1.645$, we fail to reject the null hypothesis. Therefore, we cannot conclude that the proportion of homes heated by gas has changed.

(b) p-value $= P(z \le -1.14) + P(z \ge 1.14) = 2P(z \ge 1.14) = 2(0.5 - 0.3729) = 0.2542$

Question 9. A group in favor of freezing production of nuclear weapons believes that the proportion of individuals in favor of a nuclear freeze is greater for those who have seen the movie "The Day After" (population 1) than those who have not (population 2). In an attempt to verify this belief, random samples of size 500 are obtained from the populations of interest. Among those who had seen "The Day After", 228 were in favor of a freeze. For those who had not seen the movie, 196 favored a freeze.

(a) [1] Set up the appropriate null and alternative hypotheses.
(b) [1] Set up the rejection region for this test using $\alpha = 0.05$.
(c) [2] Find the appropriate test statistic.
(d) [1] State and interpret your conclusion.
(e) [1] Find the p-value for this test.

Solution: Denote by p_1 the proportion of individuals in favor of a nuclear freeze among those who have seen the movie "The Day After", and by p_2 the proportion of individuals in favor of a nuclear freeze among those who have not seen the movie "The Day After".

(a) $H_0: \ p_1 - p_2 = 0$ vs. $H_a: \ p_1 - p_2 > 0$
(b) Reject the null hypothesis if $z > 1.645$.
(c)

$$\widehat{p}_1 = 228/500 = 0.456, \text{ and } \widehat{p}_2 = 196/500 = 0.392.$$

The pooled estimate for p required for the standard error is

$$\widehat{p} = \frac{228 + 196}{1000} = 0.424.$$

Then, the test statistics is

$$z = \frac{0.456 - 0.392}{0.0313} = 2.045.$$

 Continued on next page...

(d) Since $z > 1.645$, we reject the null hypothesis. Based on this data, the proportion in favor of a freeze that have seen the movie is greater than the proportion in favor of a freeze that have not seen the movie.

(e) p-value= $P(|Z| > 2.045) \approx 0.041$.

Question 10. The average gas mileage of a 4-wheel drive truck is 18.2 miles per gallon. The gas mileage for seven randomly selected trucks are 17.4, 18.1, 18.6, 19.2, 18.3, 18.2, and 18.0. Assume that the gas mileage distribution is normal. It is of interest to know if the sample data suggest the average gas mileage is different from 18.2 miles per gallon.

(a) [1] State the appropriate hypotheses.
(b) [2] Compute the test statistic for the hypotheses in part
(c) [1] compute the appropriate p-value associated with the test statistic in the previous question. Does the sample data support the null hypothesis at the $\alpha = 0.01$ level?

Solution: Denote by μ the average gas mileage.

(a) $H_0 : \mu = 18.2$ vs $H_a : \mu \neq 18.2$

(b) The sample mean is $\bar{x} = 18.257$, and the sample standard deviation is $s = 0.553$. Hence, the test statistic is

$$t = \frac{18.257 - 18.2}{\frac{0.553}{\sqrt{7}}} = 0.2727$$

(c) Under H_0 the test statistic T has t distribution with 6 degrees of freedom. Hence

$$p - \text{value} = 2P(Y > 0.2727) = 0.7942 \quad \text{(by using Minitab)}$$

or simply by using t-table p-value> 0.20. Since p-value exceed α, it follows that the sample data do support the null hypothesis at the $\alpha = 0.1$ level.

Question 11. [8] 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

 Continued on next page...

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? Explain.

Solution: We wish to test

$$H_0 : \mu_1 - \mu_2 = 0, \text{ vs. } H_a : \mu_1 - \mu_2 \neq 0$$

or

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

It is paired data where d_i:

$$-11, -3, 3, 14, -8, -10, -5, -7, 2, -2, -12$$

$$\bar{d} = \sum d_i/n = -39/11 = -3.55$$

$$s_d^2 : \frac{\sum d_i^2 - (\sum d_i)^2/n}{n-1} = \frac{725 - 138.27}{10} = 58.67$$

The test statistic is

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}} = \frac{-3.55 - 0}{\frac{7.66}{\sqrt{11}}} = -1.54$$

and $-t_{0.025}(10) = -2.228$, so H_0 is not rejected.