STAT 2507 C & BIT 2000 A **Assignment # 5** **Fall 2011**

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

Student # ——————————— **Your LAB Section** ——————————

<u>Due date:</u> **Monday, November 28, in class**

<u>IMPORTANT NOTES:</u>

1. **You MUST write your lab section in the space provided above (C1, C2, C3, etc.)**
2. **You MUST work by yourself NOT in groups.**
3. **Use spaces left to answer lab questions.**
4. **Total mark=20. Marks for individual questions are given in [ ].**
5. **Some answers in Lab part may vary from one student to another.**

Question 1. (*Confidence interval for a mean.*) We want to build 100 confidence intervals (CIs) with confidence level $100(1-\alpha)\% = 95\%$ for the mean μ of a Poisson distribution via the following steps:

Step 1. Generate and store in columns c6-c405 100 samples of size 400 each from a Poisson with parameter $\mu = 4$ as follows:

random 100 c6-c405;

poisson 4.

Step 2. Use columns c4 and c5 to store, respectively, the means and the standard deviations of the 100 samples generated in step 1 as follows:

rmean c6–c405 c4

rstd c6–c405 c5

Step 3. Store the lower bound and the upper bound of your 95% CIs in c2 and c3, respectively, by typing successively:

*let c2=c4−1.96*c5/20*

*let c3=c4+1.96*c5/20*

Then create a column c1 containing 1 or 0 according to whether the corresponding interval [c2 , c3] covers μ or not, by typing:

let c1=(c2 <= 4 and c3 >= 4)

Finally, sum up the entries of column c1 to find how many CIs cover the value $\mu = 4$ by typing:

tally c1

[1] a. What is the percentage of confidence intervals that contain the true value $\mu = 4$? <u>96% (±3%)</u>

[1] b. How do you compare this percentage to the confidence level 95%? <u>The two percentages are very close.</u>

Question 2. (*Hypothesis testing for μ when σ is known.)*

Imagine choosing $n = 16$ women at random from a large population and measuring their heights. Assume that the heights of the women in this population are normal with $\mu = 64$ inches and $\sigma = 3$ inches. Suppose you then test the null hypothesis $H_0:\ \mu = 64$ versus the alternative that $H_a:\ \mu \neq 64$, using $\alpha = 0.10$. Assume σ is known. Simulate the results of doing this test 20 times as follows:

random 16 c1-c20;

normal 64 3.

ztest 64 3 c1-c20

[1] a. In how many tests did you reject H_0. <u>2. (Any number ≤ 6 is acceptable.)</u> Are these rejections "correct decisions" or "incorrect decisions"? Incorrect.

[1] **b.** Are the p-values all the same for the 20 tests? No.

[1] **c.** Suppose you used $\alpha = .05$ instead of $\alpha = 0.10$. Does this change any of your decisions to reject or not? Yes/No. Should it in some cases? Yes, the number of rejections may decrease.

[1] **d.** Now assume that the population really has a mean of $\mu = 63$, instead of 64., and carry out the above 20 times, (thus, use the above minitab commands with 'normal 64 3' changed to 'normal 63 3'). Once again, using $\alpha = 0.10$ and assuming σ is known, in how many tests did you reject H_0? 8. (Any number between 2 and 15 is acceptable.) Are these rejections "correct decisions" or "incorrect decisions"? Correct.

Question 3. (*Confidence intervals for μ when σ is NOT known.*)
Suppose that $n = 9$ men are selected at random from a large population. Assume the heights of the men in this population are normal, with mean $\mu = 69$ inches and $\sigma = 3$ inches. Simulate the results of this selection 20 times and in each case find a 90% confidence interval for μ using the 'tinterval' command. The following commands may be used:

```
random 9 c1-c20;
normal 69 3.
tinterval 0.90 c1-c20
```

[1] **a.** How many of your intervals contain μ? 19. (Any number > 14 is acceptable.)

[1] **b.** Would you expect all 20 of the intervals to contain μ? No. Why? $20 \times 0.9 = 18$ intervals would be expected to contain μ.

[1] **c.** Do all the intervals have the same width? No. Why (what is the theoretical width)? $2\,t_{0.05}(8) \times s/\sqrt{9}$

[1] **d.** Suppose you took 95% intervals instead of 90%. Would they be narrower or wider? Wider.

[1] **e.** How many of your intervals contain the value 71? 12. (Any number between 5 and 16, inclusive, is acceptable.

[1] **f.** Suppose you took samples of size $n = 100$ instead of $n = 9$. Would you expect more or fewer intervals to contain 71? Fewer. What about 69? Same. What about the width of the intervals for $n = 100$: Would they be narrower or wider than for $n = 9$? Narrower.

Question 4. (*Hypothesis testing for μ when σ is NOT known*)
Repeat Question 2, using ttest instead of ztest, and answer parts (a), (b), and (c) again. (Thus 'ztest 64 3 c1-c20' changes to 'ttest 64 c1-c20'.)

[1] **a.** In how many tests did you reject H_0 3. (Any number ≤ 6 is acceptable.) Are these rejections "correct decisions" or "incorrect decision"? Incorrect.

[1] **b.** Are the p-values all the same for the 20 tests? No.

[1] **c.** Suppose you used $\alpha = .05$ instead of $\alpha = 0.10$. Does this change any of your decisions to reject or not? Yes/No. Should it in some cases? Yes, the number of rejections should decrease.

ALSO do the following questions:

[1] **1.** The manager of a university bookstore wants to estimate the mean amount spent on books in the Fall term, for all students, to within \$1.5 of the true mean with 90% probability. What is the smallest sample size n needed to achieve this. You may assume that the amounts spent on books in the Fall term vary from \$370 to \$450.

(a) 193 **(b)** 383 **(c)** 482 **(d)** 192.

Solution: A 90% margin of error of estimating μ by means of the sample mean is $z_{\alpha/2} \times \sigma/\sqrt{n}$. The manager wants to satisfy the condition

$$z_{\alpha/2} \times \sigma/\sqrt{n} \leq 1.5.$$

Hence the sample size n must be chosen to have

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{1.5}\right)^2,$$

where σ is unknown and can be estimated from the sample by $\hat{\sigma}$ =Range/4. We have $\alpha = 0.1$, $z_{\alpha/2} = 1.645$, $\hat{\sigma} = (450 - 370)/4 = 20$. Therefore n should be taken bigger than $(1.645 \times 20/1.5)^2 = 481.07$, that is, $n \geq 482$.

[1] **2.** A medication for burns carries a warning that some burn patients may have a reaction to the medication. Fifty burn patients are treated with this medication and seven show a reaction to it. Find a 98% confidence interval for the proportion of patients showing a reaction to the medication based in these results.

(a) (0.026, 0.254) **(b)** (0.044, 0.236) **(c)** (0.059, 0.221) **(d)** (0.030, 0.215)

Solution: For $n = 50$ the normal approximation can be used to find the CI. The observed value of the sample proportion is $\hat{p} = 7/50 = 0.14$ and the $\alpha/2 = 0.01$ upper percentage point of the standard normal distribution is $z_{0.01} = 2.33$. Therefore the 98% CI for p is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.14 \pm 2.33\sqrt{\frac{0.14 \times 0.86}{50}} = 0.14 \pm 0.114,$$

or $(0.026, 0.254)$. Thus we can be 98% confident that the true proportion is between 2.6% and 25.4%.

[1] **3.** 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.

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

The observed value of the test statistic and the p-value are, respectively,

(a) -1.55 and 0.1212 **(b)** 1.55 and 0.0606 **(c)** 2.42 and 0.0078 **(d)** -2.42 and 0.0156

Solution: Denote by μ the current mean revenue per invoice. The problem is to test the hypotheses $H_0: \ \mu = 6.35$ vs. $H_a: \ \mu \neq 6.35$. 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

$$\begin{aligned} \overline{x} &= 2460.40/400 = 6.161 \\ s^2 &= \frac{\sum_{i=1}^{400} x_i^2 - \frac{\sum_{i=1}^{400} x_i}{n}}{n-1} = \frac{16156.728 - 15183.168}{399} = 2.44 \\ s &= \sqrt{s^2} = \sqrt{2.44} = 1.56 \end{aligned}$$

Since $|z| = 2.42 > 1.96 = z_{0.025}$, we would reject H_0 at the 5% significance level, and conclude that the mean revenue is different from 6.35. From a normal table, p-value$=P(|Z| > 2.42) = 2P(Z < -2.42) = 0.0156$.

[1] 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. Use a 90% confidence interval to estimate the difference between the proportions of firstborn or only children in the two populations from which these samples were drawn. The 90% confidence interval is

(a) (0.058, 0.242) **(b)** (0.041, 0.260) **(c)** (0.006, 0.294) **(d)** Impossible to compute

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 $\widehat{p}_1 = 140/200 = 0.70$, and $\widehat{p}_2 = 66/120 = 0.55$. Therefore an approximate 90% confidence interval is

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

Thus we can be 90% confident that the difference between the proportions of firstborn or only children in the two populations is between 0.056 and 0.242. In repeated sampling, all intervals constructed in this manner will include the parameter $(p_1 - p_2)$ 95% of the time. Since the value $p_1 - p_2 = 0$ is not in the confidence interval, it is unlikely that $p_1 = p_2$, and we may conclude that there is a difference in the population proportions.

[1] 5. In a test of hypothesis to determine whether the proportion of heads obtained when flipping a coin differs from 0.5, the coin was flipped 100 times, and the appropriate test statistic was found to be 1.80. What is the p-value of this test?

(a) 0.9216 **(b)** 0.0718 **(c)** 0.0344 **(d)** Cannot be determined.

Solution: The test statistic for testing $H_0 : p = 0.5$ vs. $H_a : p \neq 0.5$ is

$$\frac{\hat{p} - p}{\sqrt{\hat{p}\hat{q}/n}}.$$

We know that it is nearly normally $N(0, 1)$ distributed than the sample size n is large. Therefore the p-value is approximately equal to

$$p\text{-value} = P(|Z| > 1.8) = 2P(Z < -1.8) = 2 \times 0.0359 = 0.0718.$$