

## 14.30 PROBLEM SET 8 SUGGESTED ANSWERS

TA: Tonja Bowen Bishop

### Problem 1

Suppose that the sample statistics for a random sample of 10 observations from a  $N(\mu, \sigma^2)$  population are the following:

$$\begin{aligned}\bar{X} &= 5 \\ S^2 &= 4\end{aligned}$$

a. We know that

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_9$$

Since  $n$  is not very large, we cannot assume normality, but must work with the  $t$  distribution. We need to find the number  $t_{0.025,9}$  such that

$$\Pr\left(-t_{0.025,9} \leq \sqrt{n} \frac{\bar{X} - \mu}{S} \leq t_{0.025,9}\right) = 0.95$$

From our tables we find that  $t_{0.025,9} = 2.262$ . Then,

$$\begin{aligned}\Pr\left(-2.262 \leq \sqrt{n} \frac{\bar{X} - \mu}{S} \leq 2.262\right) &= 0.95 \\ \Pr\left(\bar{X} - 2.262 \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + 2.262 \frac{S}{\sqrt{n}}\right) &= 0.95\end{aligned}$$

Our random interval is thus  $\left[\bar{X} - 2.262 \frac{S}{\sqrt{n}}, \bar{X} + 2.262 \frac{S}{\sqrt{n}}\right]$ , which for this sample is equal to  $\left[5 - 2.262 \frac{2}{\sqrt{10}}, 5 + 2.262 \frac{2}{\sqrt{10}}\right]$  or  $[3.57, 6.43]$ .

b. In general, our confidence level takes the form  $\left[\bar{X} - t_{\alpha/2,9} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,9} \frac{S}{\sqrt{n}}\right]$ , which is merely a generalization from part a. The length of the interval is

$$\begin{aligned}&\left(\bar{X} + t_{\alpha/2,9} \frac{S}{\sqrt{n}}\right) - \left(\bar{X} - t_{\alpha/2,9} \frac{S}{\sqrt{n}}\right) \\ &= 2t_{\alpha/2,9} \frac{S}{\sqrt{n}}\end{aligned}$$

Given the length,  $S$ , and  $n$ , we can solve for  $t_{\alpha/2,9}$ :

$$\begin{aligned} 3.57 &= 2t_{\alpha/2,9} \frac{2}{\sqrt{10}} \\ \frac{3.57\sqrt{10}}{4} &= t_{\alpha/2,9} \\ 2.82 &= t_{\alpha/2,9} \end{aligned}$$

Then from the table, we can see that we must choose  $\alpha = 0.02$ , so we have a 98% confidence level.

### Problem 2

a.

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}$$

We can see that this is the case because

$$\sqrt{n} \frac{\bar{X} - \mu}{S} = \frac{\sqrt{n}(\bar{X} - \mu) / \sigma}{\sqrt{(n-1)S^2 / \sigma^2 (n-1)}} = \frac{Z}{\sqrt{\chi_{n-1}^2 / (n-1)}}$$

which has a  $t$  distribution with  $n - 1$  degrees of freedom.

b.

$$n \frac{(\bar{X} - \mu)^2}{S^2} \sim F_{(1, n-1)}$$

We can derive this by

$$n \frac{(\bar{X} - \mu)^2}{S^2} = \frac{n(\bar{X} - \mu)^2 / \sigma^2 (1)}{(n-1)S^2 / \sigma^2 (n-1)} = \frac{\chi_1^2 / 1}{\chi_{n-1}^2 / (n-1)}$$

which has an  $F$  distribution with 1 and  $n - 1$  degrees of freedom.

c. The square of a statistic that has a  $t$  distribution with  $k$  degrees of freedom will have an  $F$  distribution with 1 and  $k$  degrees of freedom.

### Problem 3

Since  $\sigma^2$  is known, we can use the normal distribution rather than the  $t$  distribution.

a. To construct the confidence interval, we need to find a number  $z_{0.025}$  such that

$$\Pr \left( -z_{0.025} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_{0.025} \right) = 0.95$$

From our normal distribution table, we find that  $z_{0.025} = 1.96$ . Then, we have

$$\Pr\left(-1.96 \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq 1.96\right) = 0.95$$

$$\Pr\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Thus our confidence interval is  $\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$ , which has a length of  $2 \times 1.96 \frac{\sigma}{\sqrt{n}}$ . We want to find  $n$  such that this length is equal to  $0.01\sigma$ , so we set the values equal, cancel the  $\sigma$  on both sides, and solve for  $n$ :

$$2 \times 1.96 \frac{\sigma}{\sqrt{n}} = 0.01\sigma$$

$$\sqrt{n} = \frac{3.92}{0.01} = 392$$

$$n = 153664$$

Thus we will need quite a large sample!

b. The general formula for the length of the confidence interval is  $z_{\alpha/2} \frac{2\sigma}{\sqrt{n}}$ . Given the length,  $n$ , and  $\sigma$ , we can solve for  $z_{\alpha/2}$ , which will determine the confidence level.

$$z_{\alpha/2} \frac{2\sigma}{\sqrt{n}} = 0.01\sigma$$

$$z_{\alpha/2} = \frac{0.01}{2} \sqrt{153664/2}$$

$$= 1.39$$

We then turn to our table to see that this value corresponds to  $\alpha/2 = 0.0823$ , or a confidence level of 83.5%.

c. Again our confidence interval will be  $\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$ , and using  $\sigma^2 = 2$ , we can substitute in to get  $[\bar{X} - 0.01, \bar{X} + 0.01]$ , so the length is 0.02.

#### Problem 4

a. Notice that the landlord is constructing a confidence interval for  $X_i$ , not  $\bar{X}$ . To arrive at  $600 \pm 200$ , she first assumes that the  $X_i$  are iid normally distributed, with mean 600 and variance  $100^2$ , and that this variance is known. She then approximates the true confidence interval of  $600 \pm 1.96\sigma$  by  $600 \pm 2(100)$ . Because the confidence interval is for  $X_i$ , not  $\bar{X}$ , we cannot appeal to a CLT (and her sample size would be too small in any case), so if the  $X_i$  are not normally distributed, the confidence interval is wrong. In addition, because she is actually using an estimate of the variance, she ought to be using a  $t$  distribution with 8 degrees of freedom instead of a  $z$ , which would produce a confidence interval of  $600 \pm 2.306(100)$ . Because

the landlord is constructing a confidence interval of  $X_i$ , not  $\bar{X}$ , the size of the confidence interval will not decrease as  $n$  increases.

b. Here we will construct a 95% confidence interval for the population mean based on the assumptions that the  $X_i$  are normally distributed and the variance is known to be  $100^2$ .

$$\begin{aligned} \Pr\left(-z_{0.025} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_{0.025}\right) &= 0.95 \\ \Pr\left(\bar{X} - z_{0.025} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{0.025} \frac{\sigma}{\sqrt{n}}\right) &= 0.95 \\ \Pr\left(600 - (1.96) \frac{100}{\sqrt{9}} \leq \mu \leq 600 + (1.96) \frac{100}{\sqrt{9}}\right) &= 0.95 \\ \Pr(534.67 \leq \mu \leq 665.33) &= 0.95 \end{aligned}$$

The width of the confidence interval is  $2(1.96) \frac{\sigma}{\sqrt{n}} = \frac{392}{\sqrt{n}}$ . Thus if we want the length of the confidence interval to be less than 100, we manipulate  $\frac{392}{\sqrt{n}} < 100$  to get  $n > (3.92)^2$ , which implies  $n \geq 16$ .

### Problem 5

A confidence interval of 0.01 is very small, as we have seen before, so we will probably need a very large sample. When  $n$  is large, the sample mean, which in this case will be the number of successes divided by the total number of observations (i.e. the percentage of the sample that is a success), will be approximately normally distributed, by the central limit theorem:

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

(Recall that the variance of a Bernoulli random variable is  $p(1-p)$ ). We have seen that the length of a 95% confidence interval will in general have the form  $2(1.96) \frac{\sigma}{\sqrt{n}} = 2(1.96) \sqrt{\frac{p(1-p)}{n}}$ . The difficulty in this case is that this expression depends on  $p$ . However, you can easily show that  $p(1-p)$  is maximized at  $p = 0.5$ , so the length will be at most  $2(1.96) \sqrt{\frac{1}{4n}} = \frac{1.96}{\sqrt{n}}$ . Thus to guarantee that our confidence interval will be no more than 0.01 in width, we set  $0.01 \geq \frac{1.96}{\sqrt{n}}$ , and find that we must have  $n \geq (196)^2 = 38416$ . This number is sufficiently large to justify the use of the CLT.