

PSET 10 - Due: Sunday, July 28

1.

(a) Find the $Z_{\frac{\alpha}{2}}$ critical value for each of the following confidence levels.

(i) 99.5%

(ii) 98.5%

(iii) 97%

(b) Find the confidence level associated with the confidence interval $\bar{x} \pm 1.53 \frac{\sigma}{\sqrt{n}}$.

a.i)

$$\begin{aligned}
100(1 - \alpha)\% &= 99.5\% \\
\implies 1 - \alpha &= 0.995 \\
\implies \alpha &= 0.005 \\
\implies \frac{\alpha}{2} &= 0.0025 \\
\implies Z_{\frac{\alpha}{2}} &= Z_{0.0025} = 2.81.
\end{aligned}$$

a.ii)

$$\begin{aligned}
100(1 - \alpha)\% &= 98.5\% \\
\implies 1 - \alpha &= 0.985 \\
\implies \alpha &= 0.015 \\
\implies \frac{\alpha}{2} &= 0.0075 \\
\implies Z_{\frac{\alpha}{2}} &= Z_{0.0075} = 2.43.
\end{aligned}$$

a.iii)

$$\begin{aligned}
100(1 - \alpha)\% &= 97\% \\
\implies 1 - \alpha &= 0.97 \\
\implies \alpha &= 0.03 \\
\implies \frac{\alpha}{2} &= 0.015 \\
\implies Z_{\frac{\alpha}{2}} &= Z_{0.015} = 2.17.
\end{aligned}$$

b.)

$$\begin{aligned}
&\bar{x} \pm 1.53 \frac{\sigma}{\sqrt{n}} \\
\implies Z_{\frac{\alpha}{2}} &= 1.53 \\
\implies \frac{\alpha}{2} &= 0.0630 \\
\implies \alpha &= 0.126 \\
\implies 100(1 - 0.126)\% &= 87.4\%.
\end{aligned}$$

2. As part of a major health study, a random sample of 50 Canadians was obtained and the blood lead level of each person was recorded, resulting in the sample mean $\bar{x} = 1.254$. Suppose that blood lead levels for all Canadians vary according to a normal distribution with population standard deviation $\sigma = 0.3 \mu\text{g/dL}$.

- (a) Calculate a 95% confidence interval for μ .
- (b) Write an interpretation of the interval found in part (a).
- (c) Calculate a 90% confidence interval for μ .
- (d) A researcher desires a 99% confidence interval to have an overall width of no more than 0.1. What is the smallest sample size that will be needed to achieve this?

a.) A 95% confidence interval for the point estimate $\bar{x} = 1.254$, with population standard deviation $\sigma = 0.3$ is given by

$$\begin{aligned} & \bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} \\ \implies & \bar{x} - 1.96 \cdot \frac{0.3}{\sqrt{50}} < \mu < \bar{x} + 1.96 \cdot \frac{0.3}{\sqrt{50}} \\ \implies & 1.254 - 1.96 \cdot 0.0424 < \mu < 1.254 + 1.96 \cdot 0.0424 \\ \implies & 1.1709 < \mu < 1.3371. \end{aligned}$$

b.) The 95% CI (1.1709, 1.3371) implies that before the sample is acquired, we are 95% confident that the population mean μ will fall in this interval. It does **not** mean that μ has a 95% percent chance of being in the interval. The population mean μ is a fixed value, and is either in the interval or it is not. In the long run, if many samples are taken, we expect 95% of the computed intervals to contain the population parameter μ .

c.) For a 90% CI, we have

$$\begin{aligned} 100(1 - \alpha)\% &= 90\% \\ \implies \alpha &= 0.1 \\ \implies Z_{\frac{\alpha}{2}} &= Z_{0.05} = 1.645. \end{aligned}$$

Thus, we have a confidence interval

$$\begin{aligned} & \bar{x} \pm 1.645 \cdot \frac{\sigma}{\sqrt{n}} \\ \implies & 1.254 - 1.645 \cdot \frac{0.3}{\sqrt{50}} < \mu < 1.254 + 1.645 \cdot \frac{0.3}{\sqrt{50}} \\ \implies & 1.1842 < \mu < 1.3238. \end{aligned}$$

Remark. Computing n based on desired width is given by

$$n = \left\lceil \left(2z_{\alpha/2} \cdot \frac{\sigma}{w} \right)^2 \right\rceil.$$

Alternately, with $E = \frac{w}{2}$

$$n = \left\lceil \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{E} \right)^2 \right\rceil.$$

For a 99% CI, we have

$$\begin{aligned}100(1 - \alpha)\% &= 99\% \\ \implies \alpha &= 0.01 \\ \implies \frac{\alpha}{2} &= 0.005.\end{aligned}$$

Thus,

$$\begin{aligned}n &= \left(2 \cdot Z_{0.005} \cdot \frac{0.3}{0.1}\right)^2 \\ &= (2 \cdot 2.575 \cdot 3)^2 \\ \therefore n &= 239.\end{aligned}$$

3. The Boeing 747-8 has better fuel efficiency than the 747-400 and the A380, which means lower carbon emissions. Suppose the carbon emissions from all flights of a 747-8 is normally distributed with population standard deviation 8.7 (grams CO_2 per seat-km). Suppose that a random sample of twelve 747-8 test flights was obtained and the carbon emissions was measured for each, resulting in a sample mean of 72.

- (a) Calculate a 90% confidence interval to estimate the true mean carbon emissions.
- (b) Suppose that we wish for the sample mean \bar{x} to be accurate to within ± 3 of the true mean μ . How large a sample should be selected to insure this with 95% confidence?

a.) For a 90% CI, we have

$$\begin{aligned}\alpha &= 0.1 \\ \implies \frac{\alpha}{2} &= 0.05 \\ \implies Z_{\frac{\alpha}{2}} &= Z_{0.05} = 1.645.\end{aligned}$$

Thus, the CI is

$$\begin{aligned}72 - 1.645 \cdot \frac{8.7}{\sqrt{12}} &< \mu < 72 + 1.645 \cdot \frac{8.7}{\sqrt{12}} \\ \implies 67.8686 &< \mu < 76.1314.\end{aligned}$$

b.) With width $w = 6$, standard deviation $\sigma = 8.7$

$$\begin{aligned}n &= \left(2 \cdot Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{w}\right)^2 \\ &= \left(2 \cdot 1.96 \cdot \frac{8.7}{6}\right)^2 \\ \therefore n &= 129.\end{aligned}$$

4. Let μ denote the mean alcohol content for the population of all bottles of a certain brand of cough syrup. Suppose that a random sample of 50 bottles of this brand is selected and the alcohol content of each bottle is measured. Suppose that the data is used to calculate a 95% confidence interval for estimating μ and that interval ranges from 7.8 to 9.4.

- (a) Would a 90% confidence interval, calculated from the same sample, be longer or shorter than the given interval?
- (b) If the sample consisted of 75 bottles, would the 95% confidence interval based on the larger sample be longer or shorter than the interval given above?
- (c) Consider the following statement: "We are 95% confident that the interval from 7.8 to 9.4 captures the sample mean." Is this statement correct? Why or why not?
- (d) Consider the following statement: "95% of all bottles of this brand of cough syrup have an alcohol content between 7.8 and 9.4." Is this statement correct? Why or why not?

a.) Shorter, a smaller CI implies a tighter interval. In order to have a CI like 99%, we would need a large interval and therefore a large margin of error to insure with near certainty our parameter would fall in the interval. Inversly, a smaller CI like 90% would not require such an extreme margin of error, hence leading to a smaller interval.

b.) We can examine the formula for n to see what happens to the CI when n grows without bound. We know

$$n = \lceil (Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{E}) \rceil.$$

Where E is the margin of error, or half the intervals width ($\frac{w}{2}$). If we rearrange for E , we find

$$n = \left(Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{E} \right)^2$$

$$\implies E = \frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}}.$$

As $n \rightarrow \infty$, the quantity $\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}} \rightarrow 0$, which in turn implies $E \rightarrow 0$. Thus, we conclude as n grows without bound, for the same $Z_{\frac{\alpha}{2}}$ and σ , the margin of error (and therefore the width of the interval) shrinks toward zero.

Thus, increasing the sample size from 50 to 75 would lead to a shorter interval.

c.) This statement is not correct. The correct statement would regard the population mean μ . Ie "We are 95% confident that the interval from 7.8 to 9.4 captures the population mean μ ". Confidence intervals based on a sample mean build an interval that aims to capture the population parameter, not the sample statistic.

d.) This is also incorrect. A 95% confidence interval implies we are 95% confident that the population mean μ is inside the interval. It says nothing about individual bottles. The individual bottles will have alcohol content outside this range. The interval instead pertains to where the population mean is likely to fall.

5. A sample of 78 anchor bolts was randomly selected and the shear strength (in kips) of each was measured, resulting in a sample mean of 4.25 and a sample standard deviation of 1.30.

- (a) Calculate a 98% confidence interval to estimate the true mean shear strength.
- (b) Suppose that the shear strength for all bolts in the population was not normally distributed. Would the confidence interval found in part (a) still be valid? Why or why not?

Remark. If n is sufficiently large, the standardized variable

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has approximately a standard normal distribution. This implies that

$$\bar{X} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

is a **large-sample confidence interval** for μ with confidence level approximately $100(1 - \alpha)\%$. This formula is valid regardless of the shape of the population distribution. ☺

a.) For a 98% CI, we have

$$\begin{aligned} 100(1 - \alpha)\% &= 98\% \\ \implies \alpha &= 0.02 \\ \implies Z_{\frac{\alpha}{2}} &= Z_{0.01} = 2.33. \end{aligned}$$

Thus,

$$\begin{aligned} \bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} &< \mu < \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \\ \implies 4.25 - 2.33 \cdot \frac{1.3}{\sqrt{78}} &< \mu < 4.25 + 2.33 \cdot \frac{1.3}{\sqrt{78}} \\ \implies 3.907 &< \mu < 4.593. \end{aligned}$$

b.) Yes, since n is large the random variables \bar{X} and S are approximately normal.

6. It is important that face masks used by firefighters be able to withstand high temperatures because firefighters commonly work in temperatures of $200 - 500^\circ \text{ F}$. In a test of one type of mask, 11 of 55 masks had lenses pop out at 250° .

- (a) Construct a 95% confidence interval for the true proportion of masks of this type whose lenses would pop out at 250° .
- (b) (i) What conditions must be verified for the confidence interval in part (a) to be valid?
(ii) Check the conditions and indicate whether or not they are satisfied.
- (c) Write an interpretation of the interval in part (a).

Remark. Let $\tilde{p} = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}}$. Then a **confidence interval for a population proportion** p with confidence level approximately $100(1 - \alpha)\%$ is

$$\tilde{p} \pm z_{\alpha/2} \sqrt{\frac{\tilde{p}\hat{q}/n + \frac{z_{\alpha/2}^2}{4n^2}}{1 + \frac{z_{\alpha/2}^2}{n}}}$$

where $\hat{q} = 1 - \hat{p}$ and, as before, the $-$ in (7.10) corresponds to the lower confidence limit and the $+$ to the upper confidence limit.

This is often referred to as the **score CI for p** .

For n large, the expression becomes approximately

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\hat{p} \cdot \frac{\hat{q}}{n}}.$$

☺

a.) First, we find \hat{p}

$$\hat{p} = \frac{x}{n} = \frac{11}{55} = 0.2.$$

For a 95% CI, we have $\alpha = 0.05$, $Z_{\frac{\alpha}{2}} = 1.96$. If we use the approximation CI, we get

$$0.2 - 1.96 \cdot \sqrt{0.2 \cdot \frac{0.8}{55}} < p < 0.2 + 1.96 \cdot \sqrt{0.2 \cdot \frac{0.8}{55}}$$

$$0.0943 < p < 0.3057.$$

b.i) The conditions we must check are

- 1. $np \geq 10$
- 2. $n(1 - p) \geq 10$

b.ii) We see

$$np = 55 \cdot 0.2 = 11 \geq 10$$

$$n(1 - p) = 55 \cdot 0.8 = 44 \geq 10.$$

Thus, the conditions are satisfied

c.) Before the sample is taken, we are 95% confident that the population proportion p will fall within this interval. If many samples are taken, we expect 95% of the computed intervals using the sample proportion \hat{p} to contain the population proportion p

7. A state legislator wishes to survey residents of her district to see what proportion of them are aware of her position on an important issue.

- (a) Suppose the legislator desires a 95% confidence interval for p to have an overall width of at most 0.10, regardless of the value of \hat{p} . What is the smallest sample size that will be needed to achieve this?
- (b) Suppose the legislator has strong reason to believe that roughly 2/3 of the electorate know her position. Now how large a sample is necessary?
- (c) Unaware of how to calculate the necessary sample size, the legislator selects a random sample of 300 residents and finds that 180 of them are aware of her stance. Calculate the 95% confidence interval for estimating p .

Remark. The sample size needed for a population proportion confidence interval with width w is given by

$$n \approx \frac{4z^2\hat{p}\hat{q}}{w^2}.$$

If \hat{p} is unknown using $\hat{p} = 0.5$ gives the largest possible value of n ☹

a.) Thus, for a 95% CI with width $w = 0.10$, $\hat{p} = 0.5$, we have

$$n \approx \frac{4(1.96)^2 \cdot 0.5 \cdot 0.5}{0.1^2} = 384.16 = 385.$$

b.) With $\hat{p} = 0.6667$, we have

$$n \approx \frac{4(1.96)^2 \cdot 0.6667 \cdot 0.3333}{0.1^2} = 341.4585 = 342.$$

c.) For a sample size $n = 300$, with $\hat{p} = \frac{180}{300} = 0.6$, we have

$$0.6 - 1.96 \cdot \sqrt{0.6 \cdot \frac{0.4}{300}} < p < 0.6 + 1.96 \cdot \sqrt{0.6 \cdot \frac{0.4}{300}}$$

$$0.5446 < p < 0.6554.$$

8. In each of the following the sample size and the confidence level are given. Assume σ is unknown. Find the $t_{\frac{\alpha}{2}}$ critical value for the confidence interval $\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

(a) $n = 15$, 90% confidence

(b) $n = 20$, 95% confidence

(c) $n = 25$, 99% confidence

a.) For a 90% CI with $n - 1 = 15 - 1 = 14$ degrees of freedom, we have

$$\begin{aligned}\alpha = 0.1 &\implies \frac{\alpha}{2} = 0.05 \\ &\implies t_{0.05, 14} = 1.761.\end{aligned}$$

b.) For a 95% CI with $n - 1 = 20 - 1 = 19$ degrees of freedom, we have

$$\begin{aligned}\alpha = 0.05 &\implies \frac{\alpha}{2} = 0.025 \\ &\implies t_{0.025, 19} = 2.093.\end{aligned}$$

c.) for a 99% CI with $n - 1 = 25 - 1 = 24$ degrees of freedom, we have

$$\begin{aligned}\alpha = 0.01 &\implies \frac{\alpha}{2} = 0.005 \\ &\implies t_{0.005, 24} = 2.797.\end{aligned}$$

9. A random sample of 10 brands of vanilla yogurt was selected and the calorie count per serving was recorded, resulting in the following data: 130, 160, 150, 120, 120, 110, 170, 160, 110, 90.

- (a) Calculate the sample mean and the sample standard deviation.
- (b) If we want to calculate a confidence interval to estimate the true mean calorie count for the population, what must be true (or what must we assume) for the interval to be valid?
- (c) Calculate a 90% confidence interval to estimate the true mean calorie count.
- (d) Calculate a 99% confidence interval to estimate the true mean calorie count.

a.) The sample mean \bar{x} is given by

$$\begin{aligned} & \frac{1}{n} \sum x_i \\ &= \frac{1}{10} (130 + 160 + 150 + 120 + 120 + 110 + 170 + 160 + 110 + 90) \\ &= 132. \end{aligned}$$

The sample standard deviation s is given by

$$\sqrt{\frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}}.$$

Where

$$\begin{aligned} \sum x_i^2 &= (130^2 + 160^2 + 150^2 + 120^2 + 120^2 + 110^2 + 170^2 + 160^2 + 110^2 + 90^2) \\ &= 180600 \\ \left(\sum x_i\right)^2 &= (130 + 160 + 150 + 120 + 120 + 110 + 170 + 160 + 110 + 90)^2 \\ &= 1742400. \end{aligned}$$

Thus, the standard deviation is

$$\begin{aligned} s &= \sqrt{\frac{10 \cdot 180600 - 1742400}{90}} \\ &= 26.5832. \end{aligned}$$

b.) We assume that the population of interest is normal, so that X_1, \dots, X_n is a random sample from a normal distribution with μ and σ unknown

c.) A 90% CI with $\alpha = 0.1$, $\frac{\alpha}{2} = 0.05$, $t_{0.05,9} = 1.833$ is

$$\begin{aligned} & \bar{x} \pm t_{\frac{\alpha}{2}, \nu} \cdot \frac{s}{\sqrt{n}} \\ \Rightarrow & 132 \pm 1.833 \cdot \frac{26.5832}{\sqrt{10}} \\ \therefore & 116.5912 < \mu < 147.4088. \end{aligned}$$

d.) A 99% CI with $\alpha = 0.01$, $\frac{\alpha}{2} = 0.005$, and $t_{0.005,9} = 3.250$ is

$$\begin{aligned} & 132 \pm 3.250 \cdot \frac{26.5832}{\sqrt{10}} \\ \Rightarrow & 104.6794 < \mu < 159.3206. \end{aligned}$$

10. During the manufacture of certain commercial windows and doors, hot steel ingots are passed through a rolling mill and flattened to a prescribed thickness – a thickness that varies according to a normal distribution. The machinery is set to produce a steel section 0.25 inches thick. Fourteen steel sections were selected at random and the thickness of each was recorded. The data resulted in a sample mean of 0.285 inches and a sample standard deviation of 0.038 inches.

- (a) Calculate a 99% confidence interval to estimate the true mean steel section thickness.
- (b) As the rollers erode, the machine begins to produce steel sections that are too thick. Using the interval found in part (a), is there evidence to suggest that the true mean steel section thickness is more than the target value 0.25 inches? Explain why or why not.

a.) A 99% CI with $\bar{x} = 0.285$, $s = 0.038$, $\alpha = 0.01$, $\frac{\alpha}{2} = 0.005$, and $t_{0.005,13} = 3.012$ is

$$0.285 \pm 3.012 \cdot \frac{0.038}{\sqrt{14}}$$

$$\implies 0.2544 < \mu < 0.3156.$$

b.) Yes, there is evidence to suggest that the true mean steel section thickness is more than the target value 0.25 inches. By the interval computed in part a, $(0.2544 < \mu < 0.3156)$, we are 99% confident the true mean steel section thickness is contained in this interval, which has all values greater than 0.25. Thus we are very confident that the true mean is greater than the target thickness.