STM4PSD – Workshop 7 Solutions

1. **Top-left histogram:** $\mu = 15$ and $\sigma = 5$
   Top-right histogram: $\mu = 60$ and $\sigma = 1$
   Bottom-left histogram: $\mu = 30$ and $\sigma = 10$
   Bottom-right histogram: $\mu = 100$ and $\sigma = 20$

2. (a) We have, to two decimal places, SE $= \frac{s}{\sqrt{n}} = \frac{3.88}{\sqrt{43}} = 0.59$.

   (b) Using the provided result, the confidence interval is
   $$20.19 \pm 1.96 \times 0.59$$
   giving an approximate 95% confidence interval of $(19.03,\ 21.35)$.

   (c) Relative to the mean of 20.19 the interval seems pretty narrow. It tells that we can be confident that a mean reasonably close to 20 is likely.

   (d) An example is "We are highly confident that the mean design time at the New York office is somewhere in the range of 19 to 21.4 hours". Or, "We are 95% confident that the average design time is between 19 and 21.4 hours".

   (e) Looking at our confidence interval, we are very confident that the true mean is no greater than 21.4 hours. A suggestion that it is 25 hours seems exaggerated.

3. The estimated proportion is $\hat{p} = \frac{35}{149} \approx 0.235$.
   The standard error is then
   $$\text{SE} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.235(1-0.235)}{149}} \approx 0.035.$$
   This yields an approximate 95% confidence interval of
   $$\hat{p} \pm 1.96 \times \text{SE} = 0.235 \pm 1.96 \times 0.035$$
   giving an approximate 95% confidence interval of $(0.17, 0.30)$.

   So we are 95% confident that between 17% and 30% of visitors will click this link.

4. (a) $\overline{x} = \frac{1}{4}(14.3 + 20.2 + 13.5 + 17.4) = 16.35$

   (b) $s^2 = \frac{1}{4-1}\big((14.3 - 16.35)^2 + (20.2 - 16.35)^2 + (13.5 - 16.35)^2 + (17.4 - 16.35)^2\big) = \frac{1}{3}(28.25) \approx 9.42.$

   (c) SE $= \frac{s}{\sqrt{n}} = \frac{\sqrt{9.42}}{\sqrt{4}} \approx 1.53.$

   (d) A 95% confidence interval is then $16.35 \pm 1.96 \times 1.53$, giving $(13.35, 19.35)$.

5. (a) $\overline{x} = \frac{1}{6}(14.3 + 20.2 + 13.5 + 17.4 + 20.0 + 15.2) \approx 16.77$

   (b) $$\begin{aligned} s^2 &= \tfrac{1}{6-1}\big((14.3 - 16.77)^2 + (20.2 - 16.77)^2 + (13.5 - 16.77)^2 \\ &\qquad + (17.4 - 16.77)^2 + (20.0 - 16.77)^2 + (15.2 - 16.77)^2\big) \\ &= \tfrac{1}{5}(41.85) \approx 8.37 \end{aligned}$$

   (c) SE $= \frac{s}{\sqrt{n}} = \frac{\sqrt{8.37}}{\sqrt{6}} \approx 1.18.$

   (d) A 95% confidence interval is then $16.77 \pm 1.96 \times 1.18$, giving $(14.46, 19.08)$.

   (e) This is not contradictory. The confidence interval is telling us about the estimate of the *mean*; it tells us nothing about any individual outcomes that we will see.

6. (a) The sample mean is $\overline{x} = \frac{1}{6}(14.3 + 20.2 + 13.5 + 17.4 + 12.2 + 23.3) \approx 16.82$
   So the sample variance is
   $$\begin{aligned} s^2 &= \tfrac{1}{6-1}\big((14.3 - 16.82)^2 + (20.2 - 16.82)^2 + (13.5 - 16.82)^2 \\ &\qquad + (17.4 - 16.82)^2 + (12.2 - 16.82)^2 + (23.3 - 16.82)^2\big) \\ &= \tfrac{1}{5}(92.47) \approx 18.50. \end{aligned}$$

LA TROBE UNIVERSITY
All kinds of clever

This gives a standard error of

$$\text{SE} = \frac{s}{\sqrt{n}} = \frac{\sqrt{18.5}}{\sqrt{6}} \approx 1.76.$$

And hence a 95% confidence interval is

$$16.82 \pm 1.96 \times 1.76,$$

giving $(13.37, 20.27)$.

(b) You may notice that the confidence interval in this part is wider than the interval in Question 3, even though the sample size in this question is larger. This is not unusual: increasing the sample size doesn't *necessarily* reduce the size of the confidence interval, although it does make our results more trustworthy.

7. For an 80% confidence interval, we need to use $Q_Z((1+0.8)/2) = Q_Z(0.9) = 1.28$. Note that we use the same $\overline{x}$ and SE. So an 80% confidence interval is

$$\overline{x} \pm 1.28 \times \text{SE} = 16.77 \pm 1.28 \times 1.18,$$

giving $(15.26, 18.28)$.

For a 90% confidence interval, we need to use $Q_Z((1+0.9)/2) = Q_Z(0.95) = 1.64$. So a 90% confidence interval is

$$\overline{x} \pm 1.64 \times \text{SE} = 16.77 \pm 1.63 \times 1.18,$$

giving $(14.84, 18.71)$.