Group Work - Chapter 7

1 Suppose it is known that 45% of the general population believes that correlation implies causation.

(a) A survey taken of 60 students after completing a statistics course finds that 32% of them believe that correlation implies causation. Find a 95% confidence interval of the population proportion of students who have completed a statistics course who believe that correlation implies causation. Write it in both interval notation, (L, U) or $L < p < U$, and in $\hat{p} \pm ME$ notation.

32% of 60 is $0.32 \times 60 = 19.2 \Rightarrow 19$

Using StatCrunch (Proportion Stats $\rightarrow$ One Sample $\rightarrow$ With Summary) to find 95% confidence interval for 19 successes out of 60 observations:

$$L = 0.199, \ U = 0.434, \qquad \hat{p} = .3167, \qquad ME = \frac{U - L}{2} = \frac{0.434 - 0.199}{2} = 0.1175$$

95% confidence inteval =
(0.199, 0.434) or 0.199 < p < 0.434
0.3167 ± 0.1175

(b) Is the proportion of students who have completed a statistics course who believe that correlation implies causation different than the general population at a $\alpha = 0.05$ significance level? Did the statistics courses cause the difference, if it exists?

The 95% confidence interval does not contain the population value of 0.45. Thus, at a $\alpha = 0.05$ level, we have evidence that the statistics students are different from the general population in this regard.

No! We do not have evidence that the statistics course caused the difference. Correlation is not causation.

(c) If we wanted to known the population proportion of students who have completed a statistics course who believe that correlation implies causation within plus/minus 3% with 90% confidence, what sample size would be needed for a survey? Calculate with both unknown sample proportion and with proportion found in the earlier survey in part (a).

In StatCrunch (Proportion Stats $\rightarrow$ One Sample $\rightarrow$ Width/Sample Size):
Confidence level = 0.9, width = 0.06

Target proportion (unknown) = 0.5 : $n = 752$
Target proportion (survey) = 0.3167 : $n = 651$

2 Suppose winter daily maximum temperatures in Minnesota are known to be normally distributed with a standard deviation 12.5 ˚F.

(a) A random sample of 16 winter day maximum temperatures has a sample mean of 14.3 ˚F. What is 90% confidence interval for the population mean maximum temperature? Write it in both interval notation, (L, U) or $L < \mu < U$, and in $\bar{x} \pm ME$ notation.

Using StatCrunch (Z Stats → One Sample → With Summary) to find 90% confidence interval for mean 14.3, standard deviation 12.5 and sample size 16:

$$L = 9.16, \; U = 19.44, \qquad \bar{x} = 14.3, \qquad ME = \frac{U - L}{2} = \frac{19.44 - 9.16}{2} = 5.14$$

90% confidence inteval =
$(9.16, 19.44)$ or $9.16 < \mu < 19.44$
14.3 ± 5.14

(b) Suppose the file "max_temps_dec17.csv" represents a random sample of max temperatures in December. What is a 99% confidence interval based on this sample? (Don't have to write both forms.) Is the mean maximum December temperature different than the mean max winter temperature 14.3 ˚F as found in part (a) at a significance level of $\alpha = 0.01$?

Using StatCrunch (T Stats → One Sample → With Data):

$99\% \, CI = (16.79, \; 32.50)$

The 99% confidence interval does not contain the previous estimated mean of 14.3. Thus, at a $\alpha = 0.01$ level, we have evidence that the December temperatures are different from the general winter temperatures.

(c) If we wanted to find the mean winter maximum temperature within plus/minus 1.5 ˚F with 95% confidence, how many days would need to be sampled?

In StatCrunch (Z Stats → One Sample → Width/Sample Size):
Confidence level = 0.95, width = 3, Std. dev. = 12.5:

$n = 267$