

# Chapter 5 Textbook exercises

Solutions to even-numbered questions  
Statistics and statistical programming  
Northwestern University  
MTS 525

Aaron Shaw

October 15, 2020

All exercises taken from the *OpenIntro Statistics* textbook, 4<sup>th</sup> edition, Chapter 5.

## 5.4 Unexpected expenses

- (a) Adults in the United States.
- (b) The proportion of adults in the US who could not cover a \$400 expense without borrowing money or going into debt.

(c)

$$\hat{p} = \frac{322}{765} = 0.421$$

(d) The standard error ( $SE$ ).

(e) The formula for the standard error of a proportion can be used to do this:

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= \sqrt{\frac{0.421(1-0.421)}{765}} \\ &= 0.0179 \end{aligned}$$

- (f) The standard error of a point estimate is analogous to a standard deviation of a distribution of a random variable, so the answer to this question is best understood in relation to the number of standard error units between the point estimate (42%) and the news pundit's baseline expectation (50%). Since the difference is  $0.5 - 0.42 = 0.08$  and that is more than four times the standard error (0.0179 from part e above), the news pundit should be quite surprised.
- (g) Note that this concerns the distinction between  $\hat{p}$  and  $p$ . In this case, the two values are very close (0.42 vs. 0.40). The standard error does not change much:

$$\begin{aligned} SE &= \sqrt{\frac{p(1-p)}{n}} \\ &= \sqrt{\frac{0.40(1-0.40)}{765}} \\ &= 0.0177 \end{aligned}$$

## 5.8 Twitter users and news, Part I

The general formula for a confidence interval is  $point\ estimate \pm z^* \times SE$ . Where  $z^*$  corresponds to the z-score for the desired value of  $\alpha$ .

To estimate the interval from the data described in the question, identify the three different values. The point estimate is 45%,  $z^* = 2.58$  for a 99% confidence level (that's the number of standard deviations around the mean that ensure that 99% of a Z-score distribution is included), and  $SE = 2.4\%$ . With this we can plug and chug:

$$52\% \pm 2.58 \times 2.4\%$$

And that yields:

$$95\%CI = (45.8\%, 58.2\%)$$

Which means that from this data we are 99% confident that between 45.8% and 58.2% U.S. adult Twitter users get some news through the site.

## 5.10 Twitter users and news, Part II

- (a) False. See the answer to exercise 5.8 above. With  $\alpha = 0.01$ , we can consult the 99% confidence interval. It includes 50% but also goes lower. A null hypothesis of  $p = 0.50$  would not be rejected at this level.
- (b) False. The standard error of the sample proportion does not contain any information about the proportion of the population included in the sample. It estimates the variability of the sample proportion.
- (c) False. All else being equal, increasing the sample size will decrease the standard error. Consider the general formula for a standard error:  $\frac{\sigma}{\sqrt{n}}$  or the formula for the standard error of a proportion:  $\sqrt{\frac{p(1-p)}{n}}$ . A smaller value of  $n$  will result in a larger standard error.
- (d) False. All else being equal, a lower/smaller confidence interval will cover a narrower range. A higher/larger interval will cover a wider range. To confirm this, revisit the formula from the previous exercise and plug in the corresponding alpha value of .9, resulting in a  $z^*$  value of 1.28 (see the Z-score table in the back of *OpenIntro* and/or calculate this directly with the R command `qnorm(0.9)`).

## 5.17 Online communication

Key points here: (1) The hypotheses should be about the population proportion ( $p$ ), not the sample proportion. (2) The null hypothesis should have an equal sign. (3) The alternative hypothesis should have a not-equals sign and reference the null value rather than the observed sample proportion.

The correct way to set up these hypotheses is:

$$H_0 : p = 0.6$$

$$H_A : p \neq 0.6$$

## 5.30 True or false

- (a) True. See 5.10 part d above.
- (b) False. The alpha value (significance level) *is* the probability of Type 1 Error, so reducing the one reduces the other.
- (c) False. Failure to reject the null ( $H_0$ ) is evidence that we cannot conclude that the true value is different from the null. This is **very** different from evidence that the null hypothesis is true.

- (d) True. We'll revisit this in a moment below, but consider the relationship between a statistical test, the standard error, and the sample size as a sample size grows infinitely large. Given the formula for a standard error, the standard error of arbitrarily large samples approaches zero, resulting in arbitrarily precise point estimates that will result in rejecting the null hypothesis for *any* value of a test statistic for any critical value of  $\alpha$ .

### 5.35 Practical vs. statistical significance

True. If the sample size gets ever larger, then the standard error will become ever smaller. Eventually, when the sample size is large enough and the standard error is tiny, we can find statistically significant yet very small differences between the null value and point estimate (assuming they are not exactly equal).

### 5.36 Same observation, different sample size

As the sample size increases the standard error will decrease, the sample statistic (a Z-score comparing the point estimate against the null hypothesis in all of the examples developed in this chapter) will increase, and the resulting p-value will decrease.