OpenIntro Statistics

Chapter 5 Exercises

Jai Jeffryes

4/29/2020 - 5/11/2020

Notes

Point estimates and sampling variability

- Statistical inference is primarily concerned with understanding and quantifying the uncertainty of parameter estimates.
- Confidence interval.
- Hypothesis testing framework.
- Parameter of interest: entire population response.
- Point estimate: statistic obtained from sampling. A sample proportion is called $\hat{p}$.
- Error: difference between the point estimate and the parameter.
- Sampling error: Also called samping uncertainty, describes variability of an estimate between samples. Much of statistics consists of quantifying sampling error.
- Bias: describes a systematic tendency to under- or over-estimate the population value.
- In real-world applications, we never don't observe the sampling distribution, but understanding the hypothetical distribution helps us characterize and make sense of the point estimates we do observe.
- Central Limit Theorem.
  - Foundational for much of statistics.
  - In order for CLT to hold, requirements:
    - Independent observations,
    - Sample size must be sufficiently large. The typical success/failure condition is $np \geq 10$ and $n(1-p) \geq 10$.
    - Sample proportion $\hat{p}$ tests to follow a normal distribution with
      - $\mu_{\hat{p}} = p$
      - $SE^{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- Another requirement for samples sometimes is that they do not exceed 10% of the population. Otherwise, we overestimate the sampling error.
  - Finite population correction factor. $\sqrt{\frac{N-n}{n-1}}$
  - Standard error when sample > 0.10 times population size. $\sqrt{\frac{N-n}{n-1}}\sqrt{\frac{p(1-p)}{n}}$
- Substituting $\hat{p}$ for p. Since we don't know the population proportion we typically substitute $\hat{p}$ into success/failure conditions and calculating SE (the plug-in principle).
- More on CLT.
  - If np or $n(1-p)$ is small, distribution appears more discrete.
  - When smaller than 10, more skew.
  - When larger, more normal, and less discrete.
  - Center always at population proportion p. At different n's, $\hat{p}$ remains unbiased.
  - Variability decreases as sample size increases.
  - SE is largest when $p = 0.5$

Confidence intervals

- When stating an estimate for the population proportion, better to supply a plausible range of values.
- A confidence interval is like fishing with a net, expressing a range of plausible values likely to contain the population parameter.
- In a normal distribution, 95% of the data lie within 1.96 standard deviations from the mean.
- When the distribution of a point estimate qualifies for the Central Limit Theorem and therefore closely follows a normal distribution, we can construct a 95% confidence interval as: $\textit{point estimate} \pm 1.96 \times SE$, $\hat{p} = 1.96 \times \sqrt{\frac{p(1-p)}{n}}$
- It's common to round to the nearest percentage point or nearest 10th of a percent when reporting a confidence interval.
- There are three components to the confidence interval: the point estimate, z^*, and the standard error.
- Margin of error: in a confidence interval, $z^* \times SE$.
- Confidence interval for a single proportion.
  - Prepare. Identify $\hat{p}$, n, and determine α.
  - Check conditions. Verify distribution of $\hat{p}$ nearly normal. For one-proportion confidence intervals, use $\hat{p}$ in place of p.
  - Calculate. If conditions succeed, compute SE using $\hat{p}$, find z^*, and construct the interval.
  - Conclude. Interpret in context. Interpretation always makes statements about the population parameter and avoids language about the confidence interval capturing the population parameter with a certain probability.

Hypothesis testing for a proportion

- H_0, H_A
- Though we may fail to reject H_0, we do not accept it as true.
- Null value: The value we compare the parameter to.

Testing hypotheses using confidence intervals

- We can construct a confidence interval around the point estimate to see if it differs from a null value. Check conditions, calculate the standard error ($SE_{\hat{p}=\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$), pick z^* and construct the interval. If the interval includes the null value, you cannot reject the null hypothesis.
  - However, that method is imprecise. The standard error is often computed slightly differently in the context of a hypothesis test for a proportion. (I don't remember how. Pay attention when that comes up in a little bit.)

Decision errors

- Type 1 and 2. Think in terms of truth. The error is when you fail to detect the truth. Type 1, null is true, Type 2, alternative is true. Type 1 you reject the null hypothesis, when it's true. Type 2 you fail to reject the null hypothesis even the the alternative is true.
- Signficance level: α. The proportion of times we incorrectly reject the null hypotheses, often taken to be 0.05.

Formal testing using p-values

- P-value: the probability of observing data at least as favorable to the alternative hypothesis as our current dataset if the null hypothesis were true.

- When evaluating hypotheses for proportions using the p-value method, we modify checking conditions and computation of the standard error for the single proportion case in contrast to confidence intervals.
  - Check conditions using the null value p_0 instead of $\hat{p}$.
  - Calculate standard error using the null value instead of $\hat{p}$.
- When using the p-value method to evaluate a hypothesis test, we check conditions and construct the standard error using p_0, not the sample proportion. In this hypothesis test, we begin by assuming the null hypothesis is true, a different mindset from computing a confidence interval. Therefore, we use p_0 instead of $\hat{p}$.
- Compare the p-value to α to evaluate H_0.

Ex. 5.1

Identify the parameter, Part I. For each of the following situations, state whether the parameter of interest is a mean or a proportion. It may be helpful to examine whether individual responses are numerical or categorical.

- In a survey, one hundred college students are asked how many hours per week they spend on the Internet. Mean.
- In a survey, one hundred college students are asked: "What percentage of the time you spend on the Internet is part of your course work?" Mean.
- In a survey, one hundred college students are asked whether or not they cited information from Wikipedia in their papers. Proportion.
- In a survey, one hundred college students are asked what percentage of their total weekly spending is on alcoholic beverages. Mean.
- In a sample of one hundred recent college graduates, it is found that 85 percent expect to get a job within one year of their graduation date. Proportion.
- Hypothesis testing for a single proportion.
  - (Draw picture). Draw the null distribution and the tail areas of interest for computing the p-value.
  - Prepare. Identify parameter of interest, list hypotheses, determine the significance level, and identify $\hat{p}$ and n.
  - Check. Verify conditions to make sure $\hat{p}$ is nearly normal under the null distribution. For one-proportion hypothesis tests use null value to check success/failure conditions.
  - Calculate. If conditions hold, calculate standard error (again using p_0), Z-score, and identify the p-value.
  - Conclude. Evaluate the hypothesis test by comparing p-value to α, and provide a conclusion in the context of the problem.
- Choosing a significance level.
  - If making a Type 1 error is dangerous or costly, be stricter about rejecting the null hypothesis. Choose a small significance level.
  - If a Type 2 error is relatively dangerous or more costly than Type 1 errors, might be less strict about rejecting the null hypothesis by using a higher significance level.
- Statistical significance vs. practical significance.
  - If you take many samples, you'll be able to detect the slightest difference, but it might not have practical value. Statistical significance, but no practical significance.
  - One data science role, planning the size of a study. Get SMEs to identify the smallest meaningful difference from the null value. Then get a rough estimate of the population proportion p in order to estimate standard error. Then it would be possible to suggest a sample size sufficiently large to detect meaningful differences.
- One-sided hypothesis test. Value only in detecting if the population parameter is exclusively less than p_0 or exclusively higher. We don't double the tail value to determine the p-value

Ex. 5.2

Identify the parameter, Part II. For each of the following situations, state whether the parameter of interest is a mean or a proportion.

- A poll shows that 64% of Americans personally worry a great deal about federal spending and the budget deficit. Proportion.
- A survey reports that local TV news has shown a 17% increase in revenue within a two year period while newspaper revenues decreased by 6.4% during this time period. Proportions.
- In a survey, high school and college students are asked whether or not they use geolocation services on their smart phones. Proportion.
- In a survey, smart phone users are asked whether or not they use a web-based taxi service. Proportion.
- In a survey, smart phone users are asked how many times they used a web-based taxi service over the last year. Mean.

Ex. 5.3

As part of a quality control process for computer chips, an engineer at a factory randomly samples 212 chips during a week of production to test the current rate of chips with severe defects. She finds that 27 of the chips are defective.

- What population is under consideration in the data set?
  - Computer chips from one week of production at a specific factory.
- What parameter is being estimated?
  - Proportion of chips that are defective.
- What is the point estimate for the parameter?
  - 27/212 = 12.7358491%.
- What is the name of the statistic we use to measure the uncertainty of the point estimate?
  - Sampling error. **ANSWER: No, standard error.**
- Compute the value from part (d) for this context.

```
p <- 27 / 212
se <- sqrt((p * (1 - p) / 212))
```

- 0.0228962
- The historical rate of defects is 10%. Should the engineer be surprised by the observed rate of defects during the current week?
  - The observation lies 1.1948907 standard errors from the historical mean, so no.
- Suppose the true population value was found to be 10%. If we use this proportion to recompute the value in part (c) using $p = 0.1$ instead of $\hat{p}$, does the resulting value change much?

```
p <- 0.10
se <- sqrt((p * (1 - p) / 212))
```

- 0.0206041. Pretty close.

Ex. 5.4

Unexpected expense. In a random sample 765 adults in the United States, 322 say they could not cover a $400 unexpected expense without borrowing money or going into debt.

- What population is under consideration in the data set?
  - Adults in the U.S.
- What parameter is being estimated?
  - Proportion who could not cover an unexpected $400 expense.
- What is the point estimate for the parameter?
  - 0.420915.
- What is the name of the statistic we can use to measure the uncertainty of the point estimate?
  - Standard error.
- Compute the value from part (d) for this context.

```r
p <- 322 / 765 # Plug in
n <- 765
se <- sqrt((p * (1 - p)) / n)
```

- 0.01785
- A cable news pundit thinks the value is actually 50%. Should she be surprised by the data?

```r
Z <- (p - 0.50) / se
```

- Yes, the data lie -4.4305362 standard errors from her presumed value.
- Suppose the true population value was found to be 40%. If we use this proportion to recompute the value in part (e) using $p = 0.4$ instead of $\hat{p}$, does the resulting value change much?

```r
p.2 <- 0.40 # Plug in
n <- 765
se.2 <- sqrt((p.2 * (1 - p.2)) / n)
```

- Standard error 1: 0.01785
- Standard error 2: 0.0177123
- The second calculation differs from the first by -0.0001377, a difference of -0.77%.

Ex. 5.5

Repeated water samples. A nonprofit wants to understand the fraction of households that have elevated levels of lead in their drinking water. They expect at least 5% of homes will have elevated levels of lead, but not more than about 30%. They randomly sample 800 homes and work with the owners to retrieve water samples, and they compute the fraction of these homes with elevated lead levels. They repeat this 1,000 times and build a distribution of sample proportions.

- What is this distribution called?
  - Sampling distribution.

- Would you expect the shape of this distribution to be symmetric, right skewed, or left skewed? Explain your reasoning.
  - Some skew to the right will be present, but it will not be very evident. Proportions less than 0.5 skew to the right, more than 0.5, to the left. In this case, the sample size is large enough that the skew won't be very apparent.
- If the proportions are distributed around 8%, what is the variability of the distribution?

```
se.1 <- sqrt((0.08 * (1 - 0.08)) / 800)
```

- 0.0095917
- What is the formal name of the value you computed in~(c)?
  - Standard error.
- Suppose the researchers' budget is reduced, and they are only able to collect 250 observations per sample, but they can still collect 1,000 samples. They build a new distribution of sample proportions. How will the variability of this new distribution compare to the variability of the distribution when each sample contained 800 observations?

```
se.2 <- sqrt((0.08 * (1 - 0.08)) / 250)
se_diff <- se.2 - se.1
se_diff_prop <- se_diff / se.1
```

- First SE: 0.0095917
- Second SE: 0.0171581
- SE difference: 0.0075664
- SE difference (percent): 78.89%. Variability is greater with fewer samples.

Ex. 5.6

Repeated student samples. Of all freshman at a large college, 16% made the dean's list in the current year. As part of a class project, students randomly sample 40 students and check if those students made the list. They repeat this 1,000 times and build a distribution of sample proportions.

- What is this distribution called?
  - Sampling distribution.
- Would you expect the shape of this distribution to be symmetric, right skewed, or left skewed? Explain your reasoning.
  - Right skewed. np and $n(1-p)$ are 6.4 and 33.6, the first being lower than the rule-of-thumb value of 10.
- Calculate the variability of this distribution.
  - 0.0579655
- What is the formal name of the value you computed in~(c)?
  - Standard error.
- Suppose the students decide to sample again, this time collecting 90 students per sample, and they again collect 1,000 samples. They build a new distribution of sample proportions. How will the variability of this new distribution compare to the variability of the distribution when each sample contained 40 observations?
  - The variability will be lower with more samples collected. Thus, 0.0386437.

Ex. 5.7

Chronic illness, Part I. In 2013, the Pew Research Foundation reported that "45% of U.S. adults report that they live with one or more chronic conditions". However, this value was based on a sample, so it may not be a perfect estimate for the population parameter of interest on its own. The study reported a standard error of about 1.2%, and a normal model may reasonably be used in this setting. Create a 95% confidence interval for the proportion of U.S. adults who live with one or more chronic conditions. Also interpret the confidence interval in the context of the study.

- We are 95% confident the proportion of U.S. adults who live with at least one chromic condition lies within the interval, 45% $\pm$ 2.35%. (42.6%, 47.4%).

Ex. 5.8

Twitter users and news, Part I. A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter. The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion. Construct a 99% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter, and interpret the confidence interval in context.

- We are 99% confident the proportion of adult Twitter users who get at least some news on Twitter lies within the interval, 52% $\pm$ 6.19%. (45.8%, 58.2%).

Ex. 5.9

Chronic illness, Part II. In 2013, the Pew Research Foundation reported that "45% of U.S. adults report that they live with one or more chronic conditions", and the standard error for this estimate is 1.2%. Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

- We can say with certainty that the confidence interval from Exercise 5.7 contains the true percentage of U.S. adults who suffer from a chronic illness.
  - False. Point estimates are never definite.
- If we repeated this study 1,000 times and constructed a 95% confidence interval for each study, then approximately 950 of those confidence intervals would contain the true fraction of U.S. adults who suffer from chronic illnesses.
  - True.
- The poll provides statistically significant evidence (at the $\alpha = 0.05$ level) that the percentage of U.S. adults who suffer from chronic illnesses is below 50%.
  - True. In Ex. 5.7 we computed a confidence 95% confidence interval of (42.6%, 47.4%), whose range lies below 50%.
- Since the standard error is 1.2%, only 1.2% of people in the study communicated uncertainty about their answer.
  - False. Standard error refers to the variability of sampling error, not meaning of the sample.

Ex. 5.10

Twitter users and news, Part II. A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter, and the standard error for this estimate was 2.4%. Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

- The data provide statistically significant evidence that more than half of U.S. adult Twitter users get some news through Twitter. Use a significance level of $\alpha = 0.01$.
  - We are 99% confident that the percentage who get news on Twitter lies in the range (0.45808, 0.58192). Since the interval does not lie entirely above 0.50, the assessment is false.
- Since the standard error is 2.4%, we can conclude that 97.6% of all U.S. adult Twitter users were included in the study.
  - False. Though standard error is influenced by sample size, it isn't specifically a calculation of sample size.
- If we want to reduce the standard error of the estimate, we should collect less data.
  - False. Sample size is in the denominator of the standard error calculation. Reduction of the standard error requires more samples.
- If we construct a 90% confidence interval for the percentage of U.S. adults Twitter users who get some news through Twitter, this confidence interval will be wider than a corresponding 99% confidence interval.
  - True.

Ex. 5.11

Waiting at an ER, Part I. A hospital administrator hoping to improve wait times decides to estimate the average emergency room waiting time at her hospital. She collects a simple random sample of 64 patients and determines the time (in minutes) between when they checked in to the ER until they were first seen by a doctor. A 95% confidence interval based on this sample is (128 minutes, 147 minutes), which is based on the normal model for the mean. Determine whether the following statements are true or false, and explain your reasoning.

- We are 95% confident that the average waiting time of these 64 emergency room patients is between 128 and 147 minutes.
  - False. We are 100% confident about the sample statistic.
- We are 95% confident that the average waiting time of all patients at this hospital's emergency room is between 128 and 147 minutes.
  - True. Definition of confidence interval.
- 95% of random samples have a sample mean between 128 and 147 minutes.
  - False.
- A 99% confidence interval would be narrower than the 95% confidence interval since we need to be more sure of our estimate.
  - False. Greater confidence requires a wider interval.
- The margin of error is 9.5 and the sample mean is 137.5.
  - True. These measrues can be derived from the provided confidence interval. The margin of error is 9.5 and the sample mean is 137.5.

- In order to decrease the margin of error of a 95% confidence interval to half of what it is now, we would need to double the sample size.
  - False. Standard error is computed by $\sqrt{npq}$ To double it, n would have to be increased to 4 times its value.

Ex. 5.12

Mental health. Write the null and alternative hypotheses in words and using symbols for each of the following situations.

- Since 2008, chain restaurants in California have been required to display calorie counts of each menu item. Prior to menus displaying calorie counts, the average calorie intake of diners at a restaurant was 1100 calories. After calorie counts started to be displayed on menus, a nutritionist collected data on the number of calories consumed at this restaurant from a random sample of diners. Do these data provide convincing evidence of a difference in the average calorie intake of a diners at this restaurant?
  - H_0: The change in the number of calories consumed was not caused by the change in menu, but rather due to chance.
  - $H_A =$ An observed change in calorie consumption is caused by the menu changes.
- The state of Wisconsin would like to understand the fraction of its adult residents that consumed alcohol in the last year, specifically if the rate is different from the national rate of 70%. To help them answer this question, they conduct a random sample of 852 residents and ask them about their alcohol consumption.
  - H_0: If Wisconsin's rate of alcohol consumption differs from the national rate, it is due to chance.
  - H_A: Wisconsin's alcohol consumption does differ from the national rate.

Ex. 5.13

Website registration. A website is trying to increase registration for first-time visitors, exposing 1% of these visitors to a new site design. Of 752 randomly sampled visitors over a month who saw the new design, 64 registered.

- Check any conditions required for constructing a confidence interval.
  - Independence is satisfied since the samples are random.
  - $p = \frac{64}{752} = 0.0851064$. $np = 64$. $n(1-p) = 688$. Both are greater than 10.
- Compute the standard error.

```
p <- 64 / 752
se <- sqrt((p * (1 - p)) / 752)
```

- 0.0101755
- Construct and interpret a 90% confidence interval for the fraction of first-time visitors of the site who would register under the new design (assuming stable behaviors by new visitors over time).

```
me <- 1.65 * se
cinf <- p + me * c(-1, 1)
```

- 0.0683167, 0.101896. We are 90% confident that, presented with the new web site design, new visitors to the site would register in the range 6.8, 10.2 percent of the time.

Ex. 5.14

Website registration. A store randomly samples 603 shoppers over the course of a year and finds that 142 of them made their visit because of a coupon they'd received in the mail. Construct a 95% confidence interval for the fraction of all shoppers during the year whose visit was because of a coupon they'd received in the mail.

- Checking conditions.
  - Since the samples are random, independence is satisfied.
  - $p = \frac{142}{603} = 0.2354892$. $np = 142$. $n(1-p) = 461$. Both are greater than 10.
- Computing.

```
p <- 142 / 603
se <- sqrt((p * (1 - p)) / 603)
me <- 1.96 * se
cinf <- p + me * c(-1, 1)
```

- The sample proportion is: $\hat{p} = 0.2354892$.
- Standard error: 0.017279
- Margin of error: 0.0338668
- Confidence interval: 0.2016224, 0.2693561

We are 95% confident that the percentage of shoppers who visited during the year because of receiving a coupon was in the range 20.2, 26.9 percent.

Ex. 5.15

Identify hypotheses, Part I. Write the null and alternative hypotheses in words and then symbols for each of the following situations.

- A tutoring company would like to understand if most students tend to improve their grades (or not) after they use their services. They sample 200 of the students who used their service in the past year and ask them if their grades have improved or declined from the previous year.
  - H_0: Grades did not improve.
  - H_A: Grades did improve.
- Employers at a firm are worried about the effect of March Madness, a basketball championship held each spring in the US, on employee productivity. They estimate that on a regular business day employees spend on average 15 minutes of company time checking personal email, making personal phone calls, etc. They also collect data on how much company time employees spend on such non-business activities during March Madness. They want to determine if these data provide convincing evidence that employee productivity changed during March Madness.
  - H_0: Productivity does not change in March.
  - H_A: It does.
- **ANSWER:** The lessons here from observing the answer given in the book is that hypotheses must refer to the measure or statistic on which we're differentiating, and if we're examining proportions, we must include a reference to the null proportion.
  - Part A.

    * H_0: $p = 0.5$. The proportion of students whose grades improved are neither in the majority or minority.
    * H_0: $p \neq 0.5$. The grades of either a majority or minority of students improved.
  - Part B.
    * H_0: $\mu = 15$. The mean non-working time during March is 15 minutes.
    * H_A: $\mu \neq 15$. The mean non-working time during March is not 15 minutes. It has risen or lowered.

Ex. 5.16

Identify hypotheses, Part II. Write the null and alternative hypotheses in words and using symbols for each of the following situations.

- Since 2008, chain restaurants in California have been required to display calorie counts of each menu item. Prior to menus displaying calorie counts, the average calorie intake of diners at a restaurant was 1100 calories. After calorie counts started to be displayed on menus, a nutritionist collected data on the number of calories consumed at this restaurant from a random sample of diners. Do these data provide convincing evidence of a difference in the average calorie intake of a diners at this restaurant?
  - H_0: $\mu = 1100$. The mean calorie count consumed is 1100.
  - H_A: $\mu \neq 1100$. The mean calorie count consumed is not 1100. It is greater or lower..
- The state of Wisconsin would like to understand the fraction of its adult residents that consumed alcohol in the last year, specifically if the rate is different from the national rate of 70%. To help them answer this question, they conduct a random sample of 852 residents and ask them about their alcohol consumption.
  - H_0: $p = 0.7$. The rate of alcohol consumption equals the national rate.
  - H_A: $p \neq 0.7$. The rate of alcohol consumption differs from the national rate.

Ex. 5.17

Online communication. A study suggests that 60% of college student spend 10 or more hours per week communicating with others online. You believe that this is incorrect and decide to collect your own sample for a hypothesis test. You randomly sample 160 students from your dorm and find that 70% spent 10 or more hours a week communicating with others online. A friend of yours, who offers to help you with the hypothesis test, comes up with the following set of hypotheses. Indicate any errors you see.

$$H_0 : \hat{p} < 0.6$$
$$H_A : \hat{p} > 0.7$$

- H_0 should state the null proportion as equaling 0.6, not less than.
- H_A should say the proportion is not equal to 0.6. The alternative should accommodate that the proportion may be less than or greater than the nul proportion. The sample proportion, 0.7, should not appear in the hypothosis, as it is an experimental finding known only following hypothesis formation.
- **ANSWER**: oh, and the hypotheses should be statments about the population proportion p, not the sample proportion, $\hat{p}$.

Ex. 5.18

Married at 25. A study suggests that the 25% of 25 year olds have gotten married. You believe that this is incorrect and decide to collect your own sample for a hypothesis test. From a random sample of 25 year olds in census data with size 776, you find that 24% of them are married. A friend of yours offers to help you with setting up the hypothesis test and comes up with the following hypotheses. Indicate any errors you see.

$$H_0 : \hat{p} = 0.24$$
$$H_A : \hat{p} \neq 0.24$$

- Corrected: H_0: $p = 0.25$ is the proportion of 25 years olds who have married. H_A: $p \neq 0.25$. The proportion of married 25 year olds is greater or less than 0.25.

Ex. 5.19

Cyberbullying rates. Teens were surveyed about cyberbullying, and 54% to 64% reported experiencing cyberbullying (95% confidence interval). Answer the following questions based on this interval.

- A newspaper claims that a majority of teens have experienced cyberbullying. Is this claim supported by the confidence interval? Explain your reasoning.
  - Yes. The interval lies entirely above 50%. It would be better if the paper also reported the confidence level.
- A researcher conjectured that 70% of teens have experienced cyberbullying. Is this claim supported by the confidence interval? Explain your reasoning.
  - No, the conjectured point estimate lies outside the confidence interval.
- Without actually calculating the interval, determine if the claim of the researcher from part (b) would be supported based on a 90% confidence interval?
  - I'm still doing a back of the envelope calculation. The margin of error is about 5%, and that is almost 2 standard deviations. 3 standard deviations would be about another 2.5%, about 67%, far short of 70%.
  - **ANSWER**: Wrong! Lower confidence is a *narrower* interval. My conclusion is correct, but my method is not. The narrower confidence interval for a confidence level of 10% would have a smaller margin of error which would be further from incluing 70% in the interval.

Ex. 5.20

Waiting at an ER, Part II. Exercise 5.11 provides a 95% confidence interval for the mean waiting time at an emergency room (ER) of (128 minutes, 147 minutes). Answer the following questions based on this interval.

- A local newspaper claims that the average waiting time at this ER exceeds 3 hours. Is this claim supported by the confidence interval? Explain your reasoning.
  - No, 180 minutes lies outside the confidence interval.
- The Dean of Medicine at this hospital claims the average wait time is 2.2 hours. Is this claim supported by the confidence interval? Explain your reasoning.
  - Yes, 132 minutes is a value that lies within the interval.
- Without actually calculating the interval, determine if the claim of the Dean from part (b) would be supported based on a 99% confidence interval?
  - Yes, this time I have the width of the confidence interval straight in my head. This is a wider interval, which would still include the value claimed by the dean.

Ex. 5.21

Minimum wage, Part I. Do a majority of US adults believe raising the minimum wage will help the economy, or is there a majority who do not believe this? A Rasmussen Reports survey of 1,000 US adults found that 42% believe it will help the economy. Conduct an appropriate hypothesis test to help answer the research question.

- H_0: $p = 0.5$, the majority does not believe the minimum wage will help, and they do not believe it will not help.
- H_A: $p \neq 0.5$. Either a majority believes it will help, or a majority believes it won't.
- Check conditions. The check passes.

```
n <- 1000
p <- 0.5

n * p
```

```
## [1] 500
```

```
n * (1 - p)
```

```
## [1] 500
```

- Compute standard error

```
se <- sqrt((p * (1 - 0.5)) / n)
se
```

```
## [1] 0.01581139
```

- Calculate the Z-score.p-value.

```
z <- (0.42 - 0.5) / se
z
```

```
## [1] -5.059644
```

- Calculate p-value.

```
p_val <- 2 * pnorm(z)
p_val
```

```
## [1] 0.0000004200394
```

- Conclude. The p-value is 0, lower than the significance level, $\alpha = 0.05$. We reject the null hypothesis, concluding that the proportion who believe that raising the minimum wage will help the economy is not 50%, and since observed value is less than 50%, we conclude those who believe this are in the minority.

Ex. 5.22

Getting enough sleep. 400 students were randomly sampled from a large university, and 289 said they did not get enough sleep. Conduct a hypothesis test to check whether this represents a statistically significant difference from 50%, and use a significance level of 0.01.

- Prepare.
  - Parameter of interest: proportion of sleepless students.
  - H_0: $p = 0.5$, sleepless students number half.
  - H_A: $p \neq 0.5$, sleepless students do not number half. They're in either the majority or minority.
  - n: 400.
  - $\hat{p}$: $\frac{289}{n} = 0.7225$
  - Significance level, α: 0.01.
- Check. Use null value.
  - np and $n(1-p)$ are the same. $np = 200$
- Calculate.

```
n <- 400
p0 <- 0.5
phat <- 289 / n
se <- sqrt((p0 * (1 - p0)) / n)
z <- (phat - p0) / se
z
```

```
## [1] 8.9
```

```
up_tail_p <- pnorm(z, lower.tail = F)
p_value <- 2 * up_tail_p
p_value
```

```
## [1] 0.000000000000000000005584669
```

Since the p_value is less than α, we reject H_0. The Z score of the observation is greater than the null value, $p_0 = 0.5), so we conclude that the sample provides convincing evidence that sleepless students are in the majority.

Ex. 5.23

Working backwards, Part I. You are given the following hypotheses:

$$H_0 : p = 0.3$$
$$H_A : p \neq 0.3$$

We know the sample size is 90. For what sample proportion would the p-value be equal to 0.05? Assume that all conditions necessary for inference are satisfied.

```r
alpha <- 0.05
n <- 90
p0 <- 0.3
z <- c(1, -1) * qnorm(alpha / 2)
se <- sqrt((p0 * (1 - p0)) / n)
phat <- z * se + p0
phat
```

```
## [1] 0.2053247 0.3946753
```

Ex. 5.24

Working backwards, Part II. You are given the following hypotheses:

$$H_0 : p = 0.9$$
$$H_A : p \neq 0.9$$

We know that the sample size is 1,429. For what sample proportion would the p-value be equal to 0.01? Assume that all conditions necessary for inference are satisfied.

```r
alpha <- 0.01
n <- 1429
p0 <- 0.9
z <- c(1, -1) * qnorm(alpha / 2)
se <- sqrt((p0 * (1 - p0)) / n)
phat <- z * se + p0
phat
```

```
## [1] 0.8795581 0.9204419
```

Ex. 5.25

Testing for Fibromyalgia. A patient named Diana was diagnosed with Fibromyalgia, a long-term syndrome of body pain, and was prescribed anti-depressants. Being the skeptic that she is, Diana didn't initially believe that anti-depressants would help her symptoms. However after a couple months of being on the medication she decides that the anti-depressants are working, because she feels like her symptoms are in fact getting better.

- Write the hypotheses in words for Diana's skeptical position when she started taking the anti-depressants.
  - H_0: Her fibromyalgia remains unchanged. Anti-depressants did not help or exacerbate the syndrome.
  - H_A: Her fibromyalgia is different after anti-depressants. It is either better or worse.
- What is a Type~1 Error in this context?
  - Rejection of the null hypothesis when it is true, saying Diana's fibromyalgia has changed when it didn't.
- What is a Type~2 Error in this context?
  - Rejection of the alternative hypothesis, or rather, not rejecting the null hypothesis, saying Diana's fibromyalgia did not change when it did.

Ex. 5.26

Which is higher? In each part below, there is a value of interest and two scenarios (I and II). For each part, report if the value of interest is larger under scenario I, scenario II, or whether the value is equal under the scenarios.

(a) The standard error of $\hat{p}$ when (I) $n = 125$ or (II) $n = 500$.

- (I) is higher because n is in the denominator. Error reduces with more samples.

(b) The margin of error of a confidence interval when the confidence level is (I) 90% or (II) 80%.

- (I) is higher. Greater confidence requires a wider net.

(c) The p-value for a Z-statistic of 2.5 calculated based on a (I) sample with $n = 500$ or based on a (II) sample with $n = 1000$.

- (I) is higher. The Z-statistic is standardized using the standard error, which is located in the denominator. There is more error with fewer samples and greater probability that an observation could be due to chance.

(d) The probability of making a Type 2 Error when the alternative hypothesis is true and the significance level is (I) 0.05 or (II) 0.10.

- (II). A Type 2 Error is committed when you fail to reject H_0 when H_A is true. Considering p values in the interval (0.01 , 0.05), in the case of $\alpha = 0.05$, you would reject H_0 while in the case of $\alpha = 0.01$ you would not. Therefore, there is a higher probability of Type 2 Errors with a smaller significance level.

Ex. 5.27

Relaxing after work. The General Social Survey asked the question: "After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?" to a random sample of 1,155 Americans. A 95% confidence interval for the mean number of hours spent relaxing or pursuing activities they enjoy was (1.38, 1.92).

- Interpret this interval in context of the data.
  - We are 95% confident that the population mean for leisure time is between 1.38 and 1.92 hours.
- Suppose another set of researchers reported a confidence interval with a larger margin of error based on the same sample of 1,155 Americans. How does their confidence level compare to the confidence level of the interval stated above?
  - A larger margin of error equates with higher confidence.
- Suppose next year a new survey asking the same question is conducted, and this time the sample size is 2,500. Assuming that the population characteristics, with respect to how much time people spend relaxing after work, have not changed much within a year. How will the margin of error of the 95% confidence interval constructed based on data from the new survey compare to the margin of error of the interval stated above?
  - More samples equates with less error. The margin of error will diminish.

Ex. 5.28

Minimum wage, Part II. In Exercise 5.21, we learned that a Rasmussen Reports survey of 1,000 US adults found that 42% believe raising the minimum wage will help the economy. Construct a 99% confidence interval for the true proportion of US adults who believe this.

```
se <- sqrt(0.42 * (1 - 0.42) / 1000)
0.42 + c(-1, 1) * abs(qnorm(0.005)) * se
```

```
## [1] 0.3797973 0.4602027
```

Ex. 5.29

Testing for food safety. A food safety inspector is called upon to investigate a restaurant with a few customer reports of poor sanitation practices. The food safety inspector uses a hypothesis testing framework to evaluate whether regulations are not being met. If he decides the restaurant is in gross violation, its license to serve food will be revoked.

- Write the hypotheses in words.
  - H_0: The restaurant meets health regulations.
  - H_A: The restaurant grossly violates health regulations.
- What is a Type~1 Error in this context?
  - The restaurant loses its license even though it does not violate health regulations.
- What is a Type~2 Error in this context?
  - The restaurant retains its license even though it violates health regulations.
- Which error is more problematic for the restaurant owner? Why?
  - A Type 1 error is more problematic for the owner because he loses his license even though he is compliant.
- Which error is more problematic for the diners? Why?
  - A Type 2 error is more problematic for diners because there is a risk of continued exposure to unhealthy food.
- As a diner, would you prefer that the food safety inspector requires strong evidence or very strong evidence of health concerns before revoking a restaurant's license? Explain your reasoning.
  - As a diner, I would prefer merely strong evidence over very strong evidence in order to decrease the likelihood of a Type 1 error.

Note: My mnemonic for this was confusing and I didn't remember it correctly. Think instead in terms of truth. Either the null hypothesis is true or the alternative hypothesis is true. A Type 1 error is failure to judge the truth of the null hypothesis. A Type 2 error is failure to judge the truth of the alternative hypothesis. The reason this is confusing to recall is because, operationally, we never act on the alternative, only on the null. When the null is true we accept it, when the alternative is true, our action is to reject the null.

Ex. 5.30

True or false. Determine if the following statements are true or false, and explain your reasoning. If false, state how it could be corrected.

- If a given value (for example, the null hypothesized value of a parameter) is within a 95% confidence interval, it will also be within a 99% confidence interval.
  - True, because higher confidence equates with acceptance of a wider margin of error.
- Decreasing the significance level (α) will increase the probability of making a Type~1 Error.
  - False. Decreasing the signicance level makes it harder to reject the null hypothesis, which is to say that it decreases the probability of making a Type 1 error. For example, going from a conviction standard of preponderance of evidence to beyond a reasonable doubt reduces the likelihood of a wrongful conviction.
- Suppose the null hypothesis is $p = 0.5$ and we fail to reject H_0. Under this scenario, the true population proportion is 0.5.
  - False. We never say that we proved the null hypothesis is true, only that we did or did not reject it.
- With large sample sizes, even small differences between the null value and the observed point estimate, a difference often called the effect size, will be identified as statistically significant.
  - True. The larger the sample size, the smaller the standard error, in which case small differences will have greater significance than if the sample size and standard error were larger.

Ex. 5.31

Unemployment and relationship problems. A USA Today/Gallup poll asked a group of unemployed and underemployed Americans if they have had major problems in their relationships with their spouse or another close family member as a result of not having a job (if unemployed) or not having a full-time job (if underemployed). 27% of the 1,145 unemployed respondents and 25% of the 675 underemployed respondents said they had major problems in relationships as a result of their employment status.

- What are the hypotheses for evaluating if the proportions of unemployed and underemployed people who had relationship problems were different?
  - H_0: The proportion of the underemployed who had relationship problems matches that of the unemployed, $p = .027$.
- The p-value for this hypothesis test is approximately 0.35. Explain what this means in context of the hypothesis test and the data.
  - There is a 35% chance that the lower proportion observed in the sample of the underemployed is due to chance.

Ex. 5.32

Nearsighted. It is believed that nearsightedness affects about 8% of all children. In a random sample of 194 children, 21 are nearsighted. Conduct a hypothesis test for the following question: do these data provide evidence that the 8% value is inaccurate?

- H_0: $p = 0.08$. The proportion of nearsighted children is 8%.
- H_A: $p \neq 0.08$. The proportion of nearsighted children differs from 8%, either higher or lower.

- Choose significance: $\alpha = 0.05$.
- Check conditions. The check passes.

```
n <- 194
p <- 0.08

n * p
```

```
## [1] 15.52
```

```
n * (1 - p)
```

```
## [1] 178.48
```

- Compute standard error

```
se <- sqrt((p * (1 - p)) / n)
se
```

```
## [1] 0.01947772
```

- Calculate the Z-score.p-value.

```
phat <- 21 / n
z <- (phat - p) / se
z
```

```
## [1] 1.450243
```

- Calculate p-value.

```
p_val <- 2 * pnorm(z, lower.tail = F)
p_val
```

```
## [1] 0.1469908
```

- Conclude. The p-value is 0.146991, higher than the significance level, $\alpha = 0.05$. The observed value is higher than the null value. Therefore, we reject the null hypothesis, concluding that the proportion of nearsighted children differs from 8% and it is higher.

Ex. 5.33

Nutrition labels. The nutrition label on a bag of potato chips says that a one ounce (28~gram) serving of potato chips has 130 calories and contains ten grams of fat, with three grams of saturated fat. A random sample of 35 bags yielded a confidence interval for the number of calories per bag of 128.2 to 139.8 calories. Is there evidence that the nutrition label does not provide an accurate measure of calories in the bags of potato chips?

- **ANSWER**: Had to go the book for this one. Because 130 is inside the confidence interval, we do not have convincing evidence that the true average is any different from what the nutrition label suggests.

Ex. 5.34

CLT for proportions. Define the term "sampling distribution" of the sample proportion, and describe how the shape, center, and spread of the sampling distribution change as the sample size increases when $p = 0.1$.

- Point estimates from a population sample likely differ from population statistics. Repeated sampling produces different results whose distribution is called the sampling distribution. The greater the number of samples, the more closely the sampling distribution matches a Gaussian distribution. At lower sample sizes, the distribution is very discrete. Its shape is right skewed. It has the most spread. As the sample size increases, the distribution appears to gain more continuity. Its center remains at 0.1, but it rises in frequency. The spread narrows.

Ex. 5.35

Practical vs. statistical significance. Determine whether the following statement is true or false, and explain your reasoning: "With large sample sizes, even small differences between the null value and the observed point estimate can be statistically significant."

- True, because standard error is smaller with higher sample sizes.

Ex. 5.36

Same observation, different sample size. Suppose you conduct a hypothesis test based on a sample where the sample size is $n = 50$, and arrive at a p-value of 0.08. You then refer back to your notes and discover that you made a careless mistake, the sample size should have been $n = 500$. Will your p-value increase, decrease, or stay the same? Explain.

- It will decrease. See answer to 5.35.

Ex. 5.37

Gender pay gap in medicine. A study examined the average pay for men and women entering the workforce as doctors for 21 different positions.

- If each gender was equally paid, then we would expect about half of those positions to have men paid more than women and women would be paid more than men in the other half of positions. Write appropriate hypotheses to test this scenario.
  - H_0: $p_{men} = p_{women} = 0.5$. Men and woman are paid the same.
  - H_A: $p_{men} \neq p_{women}$. Either men or woman are paid on overage more than the other.
- Men were, on average, paid more in 19 of those 21 positions. Complete a hypothesis test using your hypotheses from part (a).
  - Choose significance: $\alpha = 0.05$.
  - Check conditions. The check passes.

```
n <- 21
p <- 0.5

n * p
```

```
## [1] 10.5
```

```
n * (1 - p)
```

```
## [1] 10.5
```

- Compute standard error

```
se <- sqrt((p * (1 - p)) / n)
se
```

```
## [1] 0.1091089
```

- Calculate the Z-score.p-value.

```
phat <- 19 / n
z <- (phat - p) / se
z
```

```
## [1] 3.709704
```

- Calculate p-value.

```
p_val <- 2 * pnorm(z, lower.tail = F)
p_val
```

```
## [1] 0.0002075016
```

- Conclude. The p-value is 0.000208, lower than the significance level, $\alpha = 0.05$. The observed proportion of men is higher than women. Therefore, we reject the null hypothesis, concluding that the men are paid more than women.