Review of Part V – Inference for One Parameter

R5.1. Babies.

a) This is a one-sample test because the data from Australia are census data, not sample data.

b) H_0: The mean weight of newborns in the U.S. is 7.86 pounds, the same as the mean weight of Australian babies. $(\mu = 7.86)$

H_A: The mean weight of newborns in the U.S. is not the same as the mean weight of Australian babies. $(\mu \neq 7.86)$

Randomization condition: Assume that the babies at this Missouri hospital are representative of all U.S. newborns. (Given)
Nearly Normal condition: We don't have the actual data, so we cannot look at a graphical display, but since the sample is large, it is safe to proceed.

The babies in the sample had a mean weight of 7.68 pounds and a standard deviation in weight of 1.31 pounds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean weight of U.S. newborns with a Student's *t*-model, with 112 – 1 = 111 degrees of freedom, $t_{111}\left(7.86, \frac{1.31}{\sqrt{112}}\right)$.

We will perform a one-sample *t*-test: $t = \dfrac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{7.68 - 7.86}{\frac{1.31}{\sqrt{112}}} \approx -1.45.$

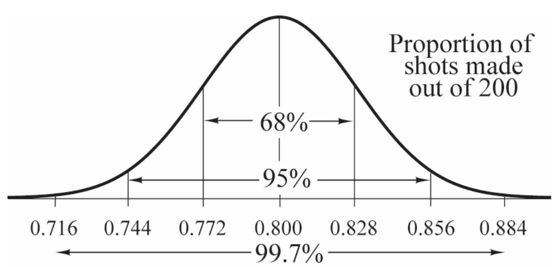
Since the P-value = 0.1487 is high, we fail to reject the null hypothesis. If we believe that the babies at this Missouri hospital are representative of all U.S. babies, there is little evidence to suggest that the mean weight of U.S. babies is different than the mean weight of Australian babies.

R5.2. Archery.

a) $p_0 = 0.80;\ SD(\hat{p}) = \sqrt{\dfrac{p_0 q_0}{n}} = \sqrt{\dfrac{(0.80)(0.20)}{200}} \approx 0.028$

b) $np = (200)(0.80) = 160$ and $nq = (200)(0.20) = 40$ are both greater than 10, so the Normal model is appropriate.

c) The Normal model of the sampling distribution of the proportion of bull's-eyes she makes out of 200 is shown below.



Approximately 68% of the time, we expect her to hit the bull's-eye on between 77.2% and 82.8% of her shots. Approximately 95% of the time, we expect her to hit the bull's-eye on between 74.4% and 85.6% of her shots. Approximately 99.7% of the time, we expect her to hit the bull's-eye on between 71.6% and 88.4% of her shots.

R5.2. (continued)

d) According to the Normal model, the probability that she hits the bull's-eye in at least 85% of her 200 shots is approximately 0.037.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.85 - 0.80}{\sqrt{\frac{(0.80)(0.20)}{200}}} \approx 1.77$$

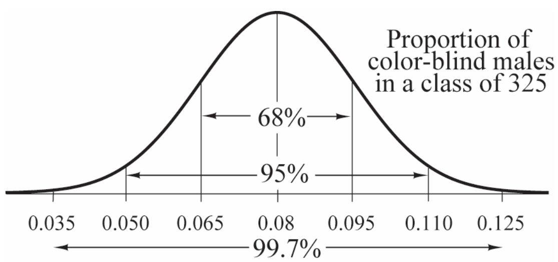
R5.3. **Color-blind.**

a) **Randomization condition:** The 325 male students are probably representative of all males.
10% condition: 325 male students are less than 10% of the population of males.
Success/Failure condition: $np = (325)(0.08) = 26$ and $nq = (325)(0.92) = 299$ are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion of colorblind men among 325 students.

b) $p_0 = 0.08;\ SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.08)(0.92)}{325}} \approx 0.015$

c)



d) According to the Normal model, we expect about 68% of classes with 325 males to have between 6.5% and 9.5% colorblind males. We expect about 95% of such classes to have between 5% and 11% colorblind males. About 99.7% of such classes are expected to have between 3.5% and 12.5% colorblind males.

R5.4. **Hamsters.**

a) **Randomization condition:** Assume these litters are representative of all litters.
Nearly Normal condition: We don't have the actual data, so we can't look at a graphical display. However, since the sample size is large, the Central Limit Theorem guarantees that the distribution of averages will be approximately Normal, as long as there are no outliers.

The litters in the sample had a mean size of 7.72 baby hamsters and a standard deviation of 2.5 baby hamsters. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's t- model, with $47 - 1 = 46$ degrees of freedom. We will use a one-sample t-interval with 90% confidence for the mean number of baby hamsters per litter.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) = 7.72 \pm t^*_{46}\left(\frac{2.5}{\sqrt{47}}\right) \approx (7.11,\ 8.33)$$

We are 90% confident that the mean number of baby hamsters per litter is between 7.11 and 8.33.

b) A 98% confidence interval would have a larger margin of error. Higher levels of confidence come at the price of less precision in the estimate.

c) We would need to sample more litters to achieve higher confidence with the same margin of error.

R5.5. Polling 2016.

a) No, the number of votes would not always be the same. We expect a certain amount of variability when sampling.

b) This is NOT a problem about confidence intervals. We already know the true proportion of voters who voted for Trump. This problem deals with the sampling distribution of that proportion.

We would expect 95% of our sample proportions of Trump voters to be within 1.960 standard deviations of the true proportion of Bush voters, 45.95%.

$$SD(\hat{p}) = \sqrt{\frac{p_T q_T}{n}} = \sqrt{\frac{(0.4595)(0.5405)}{1000}} \approx 1.576\%$$

So, we expect 95% of our sample proportions to be within 1.960(1.576%) = 3.09% of 45.95%, or between 42.86% and 49.04%.

c) Since we only expect 0.106(1000) = 10.06 votes for Jill Stein, we should be cautious when representing the sampling model with a Normal model. The predicted number of Stein votes is very close to 10. The Success/Failure condition is barely met.

d) The sample proportion of Johnson voters is expected to vary less than the sample proportion of Trump voters. Proportions farther away from 50% have smaller standard errors. We found the standard deviation of the sample proportion for Trump voters to be 1.576%. The standard deviation of the sample proportion of Johnson voters is smaller, at 0.56%.

$$SD(\hat{p}) = \sqrt{\frac{p_J q_J}{n}} = \sqrt{\frac{(0.0328)(0.9672)}{1000}} \approx 0.56\%$$

R5.6. Fake news.

a) Pew believes the true proportion of all American adults who think that fake news is causing confusion is within 3.6 percentage points of the estimated 64%—namely, between 60.4% and 67.6%.

b) **Randomization condition:** The 1002 U.S. adults were sampled randomly.
10% condition: 1002 adults are less than 10% of the population all U.S. adults.
Success/Failure condition: $n\hat{p} = (1002)(0.39) = 391$ and $n\hat{q} = (1002)(0.92) = 611$ are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion of U.S. adults who say they are "very confident" that they can recognize fake news.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.39) \pm 1.960 \sqrt{\frac{(0.39)(0.61)}{1002}} = (36.0\%, 42.0\%)$$

We are 95% confident that the interval 36.0% to 42.0% contains the true proportion of U.S. adults who would say they are "very confident" that they can recognize fake news.

c) At the same level of confidence, a confidence interval for the proportion of U.S. adults who are not at all confident in their ability to recognize fake news would be narrower than the confidence interval for the proportion of U.S. adults who are very confident in their ability to do so. The standard error for proportions farther from 0.5 is smaller than the standard error for proportions closer to 0.5. We calculated the width of the interval in the previous part to be 6 percentage points.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.06) \pm 1.960 \sqrt{\frac{(0.06)(0.94)}{1002}} = (4.5\%, 7.5\%)$$

We are 95% confident that the interval 4.5% to 7.5% contains the true proportion of U.S. adults who would say they are "not at all confident" that they can recognize fake news. This interval is only 3 percentage points wide, half the width of the previous interval.

R5.7. Scrabble.

a) The researcher believes that the true proportion of A's is within 10% of the estimated 54%, namely, between 44% and 64%.

b) A large margin of error is usually associated with a small sample, but the sample consisted of "many" hands. The margin of error is large because the standard error of the sample is large. This occurs because the true proportion of A's in a hand is close to 50%, the most difficult proportion to predict.

c) This provides no evidence that the simulation is faulty. The true proportion of A's is contained in the confidence interval. The researcher's results are consistent with 63% A's.

R5.8. Bimodal.

a) The *sample's* distribution (NOT the *sampling* distribution), is expected to look more and more like the distribution of the population, in this case, bimodal.

b) The expected value of the sample's mean is expected to be μ, the population mean, regardless of sample size.

c) The variability of the sample mean, $\sigma(\bar{y})$, is $\frac{\sigma}{\sqrt{n}}$, the population standard deviation divided by the square root of the sample size, regardless of the sample size.

d) As the sample size increases, the sampling distribution model becomes closer and closer to a Normal model.

R5.9. Gay marriage.

a) **Randomization condition:** Pew Research randomly selected 2254 U.S. adults.
10% condition: 2254 results is less than 10% of all U.S. adults.
Success/Failure condition: $n\hat{p} = (2504)(0.62) = 1552.5$ and $n\hat{q} = (2504)(0.38) = 951.5$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the percentage of U.S. adults who support marriage equality.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.62) \pm 1.960 \sqrt{\frac{(0.62)(0.38)}{2504}} = (60.1\%, 63.9\%)$$

We are 95% confident that between 66.1% and 63.9% of U.S. adults support marriage equality.

b) Since the interval is entirely above 50%, there is evidence that a majority of U.S. adults support marriage equality.

c)
$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$0.02 = 2.326 \sqrt{\frac{(0.50)(0.50)}{n}}$$
$$n = \frac{(2.326)^2 (0.50)(0.50)}{(0.02)^2}$$
$$n \approx 3382 \text{ people}$$

We do not know the true proportion of U.S. adults who support marriage equality, so use $\hat{p} = \hat{q} = 0.50$, for the most cautious estimate. In order to determine the proportion of U.S. adults who support marriage equality to within 2% with 98% confidence, we would have to sample at least 3382 people. (Using $\hat{p} = 0.62$ from the sample gives a slightly lower estimate of about 3315 people.)

R5.10. Who's the boss?

a) $P(\text{first three owned by women}) = (0.38)^3 \approx 0.055$

b) $P(\text{none of the first four are owned by women}) = (0.62)^4 \approx 0.148$

c) $P(\text{sixth firm called is owned by women} \mid \text{none of the first five were}) = 0.38$

Since the firms are chosen randomly, the fact that the first five firms were owned by men has no bearing on the ownership of the sixth firm.

R5.11. Living at home.

a) **Randomization condition:** Though not specifically stated, we will assume Pew Research randomly selected 648,118 U.S. 18-to-24-year-olds.
10% condition: 648,118 is less than 10% of all U.S. 18-to-24-year-olds.
Success/Failure condition: $n\hat{p} = (648{,}118)(0.321) = 208{,}046$ and $n\hat{q} = (648{,}118)(0.679) = 440{,}072$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of U.S. 18-to-24-year-olds who live with their parents.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.321) \pm 1.960\sqrt{\frac{(0.321)(0.679)}{648{,}118}} = (32.0\%, 32.2\%)$$

b) We are 95% confident that between 32.0% and 32.2% of U.S. 18-to-24-year-olds live with their parents.

c) 95% of all random samples of 648,118 U.S. 18-to-24-year-olds will produce confidence intervals that will capture the true proportion of U.S. 18-to-24-year-olds who live with their parents.

R5.12. Polling disclaimer.

a) It is not clear what specific question the pollster asked. Otherwise, they did a great job of identifying the W's.

b) A sample that was stratified by age, sex, region, and education was used.

c) The margin of error was 4%.

d) Since "no more than 1 time in 20 should chance variations in the sample cause the results to vary by more than 4 percentage points", the confidence level is 19/20 = 95%.

e) The subgroups had smaller sample sizes than the larger group. The standard errors in these subgroups were larger as a result, and this caused the margins of error to be larger.

f) They cautioned readers about response bias due to wording and order of the questions.

R5.13. Enough eggs?

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 1.960\sqrt{\frac{(0.75)(0.25)}{n}}$$

$$n = \frac{(1.960)^2 (0.75)(0.25)}{(0.02)^2}$$

$$n \approx 1801 \text{ hens}$$

ISA Babcock needs to collect data on about 1800 hens in order to advertise the production rate for the B300 Layer with 95% confidence with a margin of error of ± 2%.

R5.14. Largemouth bass.

a) One would expect many small fish, with a few large fish.

b) We cannot determine the probability that a largemouth bass caught from the lake weighs over 3 pounds because we don't know the exact shape of the distribution. We know that it is NOT Normal.

c) It would be quite risky to attempt to determine whether or not the mean weight of 5 fish was over 3 pounds. With a skewed distribution, a sample of size 5 is not large enough for the Central Limit Theorem to guarantee that a Normal model is appropriate to describe the distribution of the mean.

d) A sample of 60 randomly selected fish is large enough for the Central Limit Theorem to guarantee that a Normal model is appropriate to describe the sampling distribution of the mean, as long as 60 fish is less than 10% of the population of all the fish in the lake.

The mean weight is $\mu = 3.5$ pounds, with standard deviation $\sigma = 2.2$ pounds. Since the sample size is sufficiently large, we can model the sampling distribution of the mean weight of 60 fish with a Normal model, with $\mu_{\bar{y}} = 3.5$ pounds and standard deviation $\sigma(\bar{y}) = \frac{2.2}{\sqrt{60}} \approx 0.284$ pounds.

$$t = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{3 - 3.5}{\frac{2.2}{\sqrt{60}}} \approx -1.76$$

According to the Normal model, the probability that 60 randomly selected fish average more than 3 pounds is approximately 0.961.

R5.15. Cheating.

a) **Randomization condition:** The 4500 students were selected randomly.
10% condition: 4500 students is less than 10% of all students.
Success/Failure condition: $n\hat{p} = (4500)(0.74) = 3330$ and $n\hat{q} = (4500)(0.26) = 1170$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion *z*-interval to estimate the percentage of students who have cheated at least once.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.74) \pm 1.645\sqrt{\frac{(0.74)(0.26)}{4500}} = (72.9\%,\ 75.1\%)$$

b) We are 90% confident that between 72.9% and 75.1% of high school students have cheated at least once.

c) About 90% of random samples of size 4500 will produce intervals that contain the true proportion of high school students who have cheated at least once.

d) A 95% confidence interval would be wider. Greater confidence requires a larger margin of error.

R5.16. Language.

a) **Randomization condition:** 60 people were selected at random.
10% condition: The 60 people represent less than 10% of all people.
Success/Failure condition: $np = (60)(0.80) = 48$ and $nq = (60)(0.20) = 12$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of 60 randomly selected people who have left-brain language control is Normal, with $p_0 = 0.80$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.80)(0.20)}{60}} \approx 0.0516$.

R5.16. (continued)

b) $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{0.75 - 0.80}{\sqrt{\dfrac{(0.80)(0.20)}{60}}} \approx -0.968$

According to the Normal model, the probability that over 75% of these 60 people have left-brain language control is approximately 0.834.

c) If the sample had consisted of 100 people, the probability would have been higher. A larger sample results in a smaller standard deviation for the sample proportion.

d) Answers may vary. Let's consider three standard deviations below the expected proportion to be "almost certain". It would take a sample of (exactly!) 576 people to make sure that 75% would be 3 standard deviations below the expected percentage of people with left-brain language control.

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$$

$$-3 = \frac{0.75 - 0.80}{\sqrt{\dfrac{(0.80)(0.20)}{n}}}$$

$$n = \frac{(-3)^2(0.80)(0.20)}{(0.75 - 0.80)^2} = 576$$

Using round numbers for n instead of z, about 500 people in the sample would make the probability of choosing a sample with at least 75% of the people having left-brain language control is a whopping 0.997. It all depends on what "almost certain" means to you.

R5.17. **Religion 2014.**

a) **Randomization condition:** We will assume that study used a random sample of American adults.
10% condition: 35,000 is less than 10% of all American adults.
Success/Failure condition: The number of "Nones" and others in the sample, 7980 and 27,020 respectively, are both more than 10, so the sample is large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.228) \pm 1.960\sqrt{\frac{(0.228)(0.772)}{35,000}} = (22.4\%,\ 23.2\%)$$

b) We are 95% confident that the proportion of all American adults who are "Nones" is between 22.4% and 23.2%.

c) 95% of all random samples of size 35,000 will produce confidence intervals that contain the true proportion of American adults who are "Nones".

R5.18. Teen smoking 2015.

Randomization condition: Assume that the freshman class is representative of all teenagers. This may not be a reasonable assumption. There are many interlocking relationships between smoking, socioeconomic status, and college attendance. This class may not be representative of all teens with regards to smoking simply because they are in college. Be cautious with your conclusions!
10% condition: The freshman class is less than 10% of all teenagers.
Success/Failure condition: $np = (522)(0.093) = 48.546$ and $nq = (522)(0.907) = 473.454$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of 522 students who smoke is Normal, with $p_0 = 0.093$ and standard deviation $SD(\hat{p}) = \sqrt{\frac{p_0q_0}{n}} = \sqrt{\frac{(0.093)(0.907)}{522}} \approx 0.0127$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{0.1 - 0.93}{\sqrt{\frac{(0.093)(0.907)}{522}}} \approx 0.551$$

10% is about 0.551 standard deviations above the expected proportion of smokers. If the true proportion of smokers is 9.3%, the Normal model predicts that the probability that more than 10% of these students smoke is approximately 0.291.

R5.19. Alcohol abuse.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 1.645\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(1.645)^2(0.5)(0.5)}{(0.04)^2}$$

$$n \approx 423$$

The university will have to sample at least 423 students in order to estimate the proportion of students who have been drunk with in the past week to within ± 4%, with 90% confidence.

R5.20. Errors.

a) Since a treatment (the additive) is imposed, this is an experiment.

b) The company is only interested in a decrease in the percentage of cars needing repairs, so they will perform a one-sided test.

c) The independent laboratory will make a Type I error if they decide that the additive reduces the number of repairs, when it actually makes no difference in the number of repairs.

d) The independent laboratory will make a Type II error if they decide that the additive makes no difference in the number of repairs, when it actually reduces the number of repairs.

e) The additive manufacturer would consider a Type II error more serious. The lab claims that the manufacturer's product doesn't work, and it actually does.

f) Since this was a controlled experiment, the company can conclude that the additive is the reason that the cabs are running better. They should be cautious recommending it for all cars. There is evidence that the additive works well for cabs, which get heavy use. It might not be effective in cars with a different pattern of use than cabs.

g) Instead of using the additive on all cars, assign half of the cars to use the additive and half as a control and compare the proportions needing repairs.

R5.21. Safety.

a) $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (0.14) \pm 1.960\sqrt{\dfrac{(0.14)(0.86)}{814}} = (11.6\%, 16.4\%)$

We are 95% confident that between 11.6% and 16.4% of Texas children wear helmets when biking, roller skating, or skateboarding.

b) These data might not be a random sample.

c) $ME = z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

$$0.04 = 2.326\sqrt{\frac{(0.14)(0.86)}{n}}$$

$$n = \frac{(2.326)^2 (0.14)(0.86)}{(0.04)^2}$$

$$n \approx 408$$

If we use the 14% estimate obtained from the first study, the researchers will need to observe at least 408 kids in order to estimate the proportion of kids who wear helmets to within 4%, with 98% confidence.

(If you use a more cautious approach, estimating that 50% of kids wear helmets, you need a whopping 846 observations. Are you beginning to see why pilot studies are conducted?)

R5.22. Fried PCs.

a) H_0: The computer is undamaged.
H_A: The computer is damaged.

b) The biggest advantage is that all of the damaged computers will be detected, since, historically, damaged computers never pass all the tests. The disadvantage is that only 80% of undamaged computers pass all the tests. The engineers will be classifying 20% of the undamaged computers as damaged.

c) In this example, a Type I error is rejecting an undamaged computer. To allow this to happen only 5% of the time, the engineers would reject any computer that failed 3 or more tests, since 95% of the undamaged computers fail two or fewer tests.

d) The power of the test in part (c) is 20%, since only 20% of the damaged machines fail 3 or more tests.

e) By declaring computers "damaged" if they fail 2 or more tests, the engineers will be rejecting only 7% of undamaged computers. From 5% to 7% is an increase of 2% in α. Since 90% of the damaged computers fail 2 or more tests, the power of the test is now 90%, a substantial increase.

R5.23. Power.

a) Power will increase, since the variability in the sampling distribution will decrease. We are more certain of all our decisions when there is less variability.

b) Power will decrease, since we are rejecting the null hypothesis less often.

R5.24. Approval 2016.

H_0 Barack Obama's final approval rating was 66%. ($p = 0.66$)
H_A Barack Obama's final approval rating was lower than 66%. ($p > 0.66$)

Randomization condition: The adults were chosen randomly.
10% condition: 1000 adults are less than 10% of all adults.
Success/Failure condition: $np = (1000)(0.66) = 660$ and $nq = (1000)(0.34) = 340$ are both greater than 10, so the sample is large enough.

R5.24. (continued)

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.66$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.66)(0.34)}{1000}} \approx 0.015.$

The observed approval rating is $\hat{p} = 0.63.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.63 - 0.66}{\sqrt{\frac{(0.66)(0.34)}{1000}}} \approx -2.00.$

Since the P-value = 0.023 is low, we reject the null hypothesis. There is convincing evidence that President Barack Obama's final approval rating was lower than the 66% approval rating of President Bill Clinton.

R5.25. Grade inflation.

H_0: In 2000, 20%of students at the university had a GPA of at least 3.5. ($p = 0.20$)
H_A: In 2000, more than 20%of students had a GPA of at least 3.5. ($p > 0.20$)

Randomization condition: The GPAs were chosen randomly.
10% condition: 1100 GPAs are less than 10% of all GPAs.
Success/Failure condition: $np = (1100)(0.20) = 220$ and $nq = (1100)(0.80) = 880$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.20$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.20)(0.80)}{1100}} \approx 0.0121.$

The observed approval rating is $\hat{p} = 0.25.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.25 - 0.20}{\sqrt{\frac{(0.20)(0.80)}{1100}}} \approx 4.15.$

Since the P-value is less than 0.0001, which is low, we reject the null hypothesis. There is strong evidence the percentage of students whose GPAs are at least 3.5 is higher in 2000 than in 1996.

R5.26. Name recognition.

a) The company wants evidence that the athlete's name is recognized more often than 25%.

b) Type I error means that fewer than 25% of people will recognize the athlete's name, yet the company offers the athlete an endorsement contract anyway. In this case, the company is employing an athlete that doesn't fulfill their advertising needs.

Type II error means that more than 25% of people will recognize the athlete's name, but the company doesn't offer the contract to the athlete. In this case, the company is letting go of an athlete that meets their advertising needs.

c) If the company uses a 10% level of significance, the company will hire more athletes that don't have high enough name recognition for their needs. The risk of committing a Type I error is higher.

At the same level of significance, the company is less likely to lose out on athletes with high name recognition. They will commit fewer Type II errors.

R5.27. Name recognition, part II.

a) The 2% difference between the 27% name recognition in the sample, and the desired 25% name recognition may have been due to sampling error. It's possible that the actual percentage of all people who recognize the name is lower than 25%, even though the percentage in the sample of 500 people was 27%. The company just wasn't willing to take that chance. They'll give the endorsement contract to an athlete that they are convinced has better name recognition.

b) The company committed a Type II error. The null hypothesis (that only 25% of the population would recognize the athlete's name) was false, and they didn't notice.

c) The power of the test would have been higher if the athlete were more famous. It would have been difficult not to notice that an athlete had, for example, 60% name recognition if they were only looking for 25% name recognition.

R5.28. Dropouts.

Randomization condition: Assume that these subjects are representative of all anorexia nervosa patients.
10% condition: 198 is less than 10% of all patients.
Success/Failure condition: The number of dropouts, 105, and the number of subjects that remained, 93, are both greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{105}{198}\right) \pm 1.960 \sqrt{\frac{\left(\tfrac{105}{198}\right)\left(\tfrac{93}{198}\right)}{198}} = (46\%,\ 60\%)$$

We are 95% confident that between 46% and 60% of anorexia nervosa patients will drop out of treatment programs. However, this wasn't a random sample of all patients. They were assigned to treatment programs rather than choosing their own. They may have had different experiences if they were not part of an experiment.

R5.29. Women.

H_0: The percentage of businesses in the area owned by women is 36%. ($p = 0.36$)
H_A: The percentage of businesses owned by women is not 36%. ($p \neq 0.36$)

Random condition: This is a random sample of 410 businesses.
10% condition: The sample of 410 businesses is less than 10% of all businesses.
Success/Failure condition: $np = (410)(0.36) = 147.6$ and $nq = (410)(0.74) = 262.4$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.36$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.36)(0.64)}{410}} \approx 0.02371.$

The observed proportion of businesses owned by women is $\hat{p} = \frac{164}{410} = 0.4.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.40 - 0.36}{\sqrt{\frac{(0.36)(0.64)}{410}}} \approx 1.687.$

Since the P-value = 0.0915 is fairly high, we fail to reject the null hypothesis. There is little evidence that the proportion of businesses in the Denver area owned by women is any different than the national figure of 36%.

R5.30. Speeding.

a) H_0: The percentage of speeding tickets issued to black drivers is 16%, the same as the percentage of registered drivers who are black. ($p = 0.16$)

H_A: The percentage of speeding tickets issued to black drivers is greater than 16%, the percentage of registered drivers who are black. ($p > 0.16$)

Random condition: Assume that this month is representative of all months with respect to the percentage of tickets issued to black drivers.
10% condition: 324 speeding tickets are less than 10% of all tickets.
Success/Failure condition: $np = (324)(0.16) = 52$ and $nq = (324)(0.84) = 272$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.16$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.16)(0.84)}{324}} \approx 0.02037.$

The observed proportion of tickets issued is $\hat{p} = 0.25$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.25 - 0.16}{\sqrt{\frac{(0.16)(0.84)}{324}}} \approx 4.42.$

Since the P-value $= 4.96 \times 10^{-6}$ is very low, we reject the null hypothesis. There is strong evidence that the percentage of speeding tickets issued to black drivers is greater than 16%.

b) There is strong evidence of an association between the receipt of a speeding ticket and race. Black drivers appear to be issued tickets at a higher rate than expected. However, this does not prove that racial profiling exists. There may be other factors present.

c) Answers may vary. The primary statistic of interest is the percentage of black motorists on this section of the New Jersey Turnpike. For example, if 80% of drivers on this section are black, then 25% of the speeding tickets being issued to black motorists is not an usually high percentage. In fact, it is probably unusually low. On the other hand, if only 3% of the motorists on this section of the turnpike are black, then there is even more evidence that racial profiling may be occurring.

R5.31. Petitions.

a) $\frac{1772}{2000} = 0.886 = 88.6\%$ of the sample signatures were valid.

b) $\frac{250{,}000}{304{,}266} \approx 0.822 \approx 82.2\%$ of the petition signatures must be valid in order to have the initiative certified by the Elections Committee.

c) If the Elections Committee commits a Type I error, a petition would be certified when there are not enough valid signatures.

d) If the Elections Committee commits a Type II error, a valid petition is not certified.

R5.31. (continued)

e) H_0: The percentage of valid signatures is 82.2% ($p = 0.822$)
H_A: The percentage of valid signatures is greater than 82.2% ($p > 0.822$)

Random Condition: This is a simple random sample of 2000 signatures.
10% condition: The sample of 2000 signatures is less than 10% of all signatures.
Success/Failure condition: $np = (2000)(0.822) = 1644$ and $nq = (2000)(0.178) = 356$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.822$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.822)(0.178)}{2000}} \approx 0.00855.$

The observed proportion of valid signatures is $\hat{p} = \frac{1772}{2000} \approx 0.886.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.886 - 0.822}{\sqrt{\frac{(0.822)(0.178)}{2000}}} \approx 7.48.$

Since the P-value $= 3.64 \times 10^{-14}$ is low, we reject the null hypothesis. There is strong evidence that the percentage of valid signatures is greater than 82.2%. The petition should be certified.

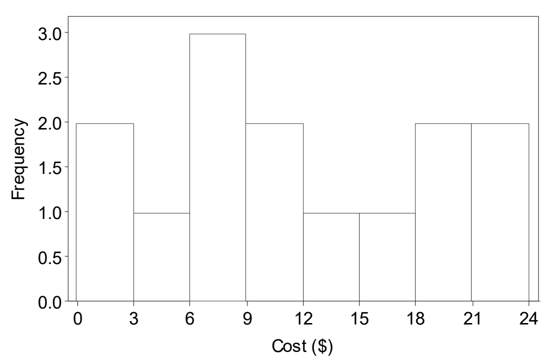
f) In order to increase the power of their test to detect valid petitions, the Elections Committee could sample more signatures.

R5.32. Meals.

H_0: The college student's mean daily food expense is \$10. $(\mu = 10)$

H_A: The college student's mean daily food expense is greater than \$10. $(\mu > 10)$

Randomization condition: Assume that these days are representative of all days.
Nearly Normal condition: The histogram of daily expenses is fairly unimodal and symmetric. It is reasonable to think that this sample came from a Normal population.



The expenses in the sample had a mean of 11.4243 dollars and a standard deviation of 8.05794 dollars. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean daily expense with a Student's t-model, with 14 – 1 = 13 degrees of freedom, $t_{13}\left(10, \frac{8.05794}{\sqrt{14}}\right)$.

R5.32. (continued)

We will perform a one-sample t-test: $t = \dfrac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{11.4243 - 10}{\frac{8.05794}{14}} \approx 0.66.$

Since the P-value = 0.2600 is high, we fail to reject the null hypothesis. There is no evidence that the student's average spending is more than \$10 per day.

R5.33. Occupy Wall Street.

Randomization condition: The 901 American adults were sampled randomly.
10% condition: 901 is less than 10% of all American adults.
Success/Failure condition: $n\hat{p} = (901)(0.599) = 540$ and $n\hat{q} = (901)(0.401) = 361$ are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{540}{901}\right) \pm 1.960 \sqrt{\frac{\left(\tfrac{540}{901}\right)\left(\tfrac{361}{901}\right)}{901}} = (56.7\%,\ 63.1\%)$$

We are 95% confident that between 56.7% and 63.1% of all American adults agree with the statement "The Occupy Wall Street protesters offered new insights on social issues."

R5.34. Streams.

Random condition: The researchers randomly selected 172 streams.
10% condition: 172 is less than 10% of all streams.
Success/Failure condition: $n\hat{p} = 69$ and $n\hat{q} = 103$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of Adirondack streams with a shale substrate.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{69}{172}\right) \pm 1.960 \sqrt{\frac{\left(\tfrac{69}{172}\right)\left(\tfrac{103}{172}\right)}{172}} = (32.8\%, 47.4\%)$$

We are 95% confident that between 32.8% and 47.4% of Adirondack streams have a shale substrate.

R5.35. Skin cancer.

a) **Independence assumption:** We must assume that the 152 patients are representative of others with skin cancer.
10% condition: 152 is less than 10% of all skin cancer patients.
Success/Failure condition: $n\hat{p} = (152)(0.53) = 81$ and $n\hat{q} = (152)(0.47) = 71$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of all skin cancer patients that would have a partial or complete response to vemurafenib.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{81}{152}\right) \pm 1.960 \sqrt{\frac{\left(\tfrac{81}{152}\right)\left(\tfrac{71}{152}\right)}{152}} = (45.36\%, 61.22\%)$$

We are 95% confident that between 45.36% and 61.22% of all patients with metastatic melanoma would have a partial or complete response to vemurafenib.

R5.35. (continued)

b) $ME = z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

$$0.06 = 1.960\sqrt{\frac{(0.53)(0.47)}{n}}$$

$$n = \frac{(1.960)^2(0.53)(0.47)}{(0.06)^2}$$

$$n \approx 266$$

If we use the 53% estimate obtained from the first study, the researchers will need to study at least 266 patients in order to estimate the proportion of skin cancer patients who would have a partial or complete response to vemurafenib to within 6% at 95% confidence. (If they use a more cautious approach with an estimate of 50%, they would need 267 patients.)

R5.36. Bread.

a) Since the histogram shows that the distribution of the number of loaves sold per day is skewed strongly to the right, we can't use the Normal model to estimate the number of loaves sold on the busiest 10% of days.

b) **Randomization condition:** Assume that these days are representative of all days.
Nearly Normal condition: The histogram is skewed strongly to the right. However, since the sample size is large, the Central Limit Theorem guarantees that the distribution of averages will be approximately Normal.

The days in the sample had a mean of 103 loaves sold and a standard deviation of 9 loaves sold. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's *t*- model, with 100 – 1 = 99 degrees of freedom. We will use a one-sample *t*-interval with 95% confidence for the mean number of loaves sold. (By hand, use $t_{50}^* \approx 2.403$ from the table.)

c) $\bar{y} \pm t_{n-1}^*\left(\dfrac{s}{\sqrt{n}}\right) = 103 \pm t_{99}^*\left(\dfrac{9}{\sqrt{100}}\right) \approx (101.2, 104.8)$

We are 95% confident that the mean number of loaves sold per day at the Clarksburg Bakery is between 101.2 and 104.8.

d) We know that in order to cut the margin of error in half, we need to a sample four times as large. If we allow a margin of error that is twice as wide, that would require a sample only one-fourth the size. In this case, our original sample is 100 loaves; so 25 loaves would be a sufficient number to estimate the mean with a margin of error twice as wide.

e) Since the interval is completely above 100 loaves, there is strong evidence that the estimate was incorrect. The evidence suggests that the mean number of loaves sold per day is greater than 100. This difference is statistically significant, but may not be practically significant. It seems like the owners made a pretty good estimate!

R5.37. Fritos®.

a) H_0: The mean weight of bags of Fritos is 35.4 grams. $(\mu = 35.4)$

H_A: The mean weight of bags of Fritos is less than 35.4 grams. $(\mu < 35.4)$

b) **Randomization condition:** It is reasonable to think that the 6 bags are representative of all bags of Fritos.
Nearly Normal condition: The histogram of bags weights shows one unusually heavy bag. Although not technically an outlier, it probably should be excluded for the purposes of the test. (We will leave it in for the preliminary test, then remove it and test again.)

R5.37. (continued)

c) The bags in the sample had a mean weight of 35.5333 grams and a standard deviation in weight of 0.450185 grams. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean weight of bags of Fritos with a Student's *t*-model, with 6 – 1 = 5 degrees of freedom, $t_5\left(35.4, \frac{0.450185}{\sqrt{6}}\right)$.

We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{35.5333 - 35.4}{\frac{0.450185}{\sqrt{6}}} \approx 0.726.$

Since the P-value = 0.7497 is high, we fail to reject the null hypothesis. There is no evidence to suggest that the mean weight of bags of Fritos is less than 35.4 grams.

d) With the one unusually high value removed, the mean weight of the 5 remaining bags is 35.36 grams, with a standard deviation in weight of 0.167332 grams. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean weight of bags of Fritos with a Student's *t*-model, with 5 – 1 = 4 degrees of freedom, $t_4\left(35.4, \frac{0.167332}{\sqrt{5}}\right)$.

We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{35.36 - 35.4}{\frac{0.167332}{\sqrt{5}}} \approx -0.53.$

Since the P-value = 0.3107 is high, we fail to reject the null hypothesis. There is no evidence to suggest that the mean weight of bags of Fritos is less than 35.4 grams.

e) Neither test provides evidence that the mean weight of bags of Fritos is less than 35.4 grams. It is reasonable to believe that the mean weight of the bags is the same as the stated weight. However, the sample sizes are very small, and the tests have very little power to detect lower mean weights. It would be a good idea to weigh more bags.

R5.38. And it means?

a) The margin of error is $\frac{(2391 - 1644)}{2} = \$373.50.$

b) The insurance agent is 95% confident that the mean loss claimed by clients after home burglaries is between \$1644 and \$2391.

c) 95% of all random samples of this size will produce intervals that contain the true mean loss claimed.

R5.39. Batteries.

a) Different samples have different means. Since this is a fairly small sample, the difference may be due to natural sampling variation. Also, we have no idea how to quantify "a lot less" without considering the variation as measured by the standard deviation.

b) H_0: The mean life of a battery is 100 hours. $(\mu = 100)$

H_A: The mean life of a battery is less than 100 hours. $(\mu < 100)$

c) **Randomization condition:** It is reasonable to think that these 16 batteries are representative of all batteries of this type
Normal population assumption: Since we don't have the actual data, we can't check a graphical display, and the sample is not large. Assume that the population of battery lifetimes is Normal.

R5.39. (continued)

d) The batteries in the sample had a mean life of 97 hours and a standard deviation of 12 hours. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean battery life with a Student's *t*-model, with 16 – 1 = 15 degrees of freedom, $t_{15}\left(100, \frac{12}{\sqrt{16}}\right)$, or $t_{15}(100, 3)$.

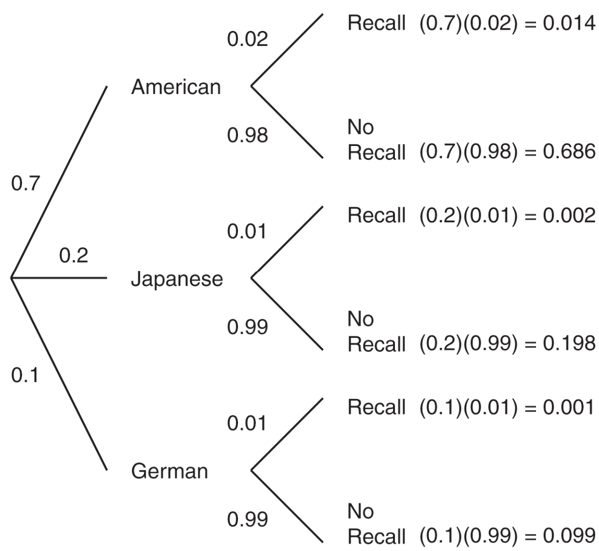
We will perform a one-sample *t*-test: $t = \dfrac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{97 - 100}{\frac{12}{\sqrt{16}}} = -1.0$

Since the P-value = 0.166 is greater than $\alpha = 0.05$, we fail to reject the null hypothesis. There is no evidence to suggest that the mean battery life is less than 100 hours.

e) If the mean life of the company's batteries is only 98 hours, then the mean life is less than 100, and the null hypothesis is false. We failed to reject a false null hypothesis, making a Type II error.

R5.40. Recalls.

Organize the information in a tree diagram.



a)
$$\begin{aligned} P(\text{recall}) &= P(\text{American recall}) + P(\text{Japanese recall}) + P(\text{German recall}) \\ &= 0.014 + 0.002 + 0.001 \\ &= 0.017 \end{aligned}$$

b)
$$P(\text{American} \mid \text{recall}) = \frac{P(\text{American and recall})}{P(\text{recall})} = \frac{0.014}{0.014 + 0.002 + 0.001} \approx 0.824$$