Chapter 18 – Testing Hypotheses

Section 18.1

1. Better than aspirin?

The new drug is not more effective than aspirin, and reduces the risk of heart attack by 44%. ($p = 0.44$)

2. Psychic.

Your friend's chance of guessing the correct suit is 25%. ($p = 0.25$)

3. Parameters and hypotheses.

a) Let $p =$ probability of winning on a slot machine.
H_0: $p = 0.01$ vs. H_A: $p \neq 0.01$

b) Let $\mu =$ mean spending per customer this year.
H_0: $\mu = \$35.32$ vs. H_A: $\mu \neq \$35.32$

c) Let $p =$ proportion of patients cured by the new drug.
H_0: $p = 0.3$ vs. H_A: $p \neq 0.3$

d) Let $p =$ proportion of clients now using the website.
H_0: $p = 0.4$ vs. H_A: $p \neq 0.4$

4. Hypotheses and parameters.

a) Let $p =$ proportion using seat belts in MA.
H_0: $p = 0.65$ vs. H_A: $p \neq 0.65$

b) Let $p =$ proportion of employees willing to pay for onsite day care.
H_0: $p = 0.45$ vs. H_A: $p \neq 0.45$

c) Let $p =$ probability of default for Gold card customers.
H_0: $p = 0.067$ vs. H_A: $p \neq 0.067$

d) Let $\mu =$ mean time (in months) that regular customers have been customers of the bank. H_0: $\mu = 17.3$ vs. H_A: $\mu \neq 17.3$

Section 18.2

5. Better than aspirin again?

a) The alternative to the null hypothesis is one-sided. They are interested in discovering only if their drug is more effective than aspirin, not if it is less effective than aspirin.

b) Since the P-value of 0.0028 is low, reject the null hypothesis. There is evidence that the new drug is more effective than aspirin.

c) Since the P-value of 0.28 is high, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the new drug is more effective than aspirin.

6. GRE performance.

a) The alternative to the null hypothesis is one-sided. They claim that more than 50% of students who take their course improve their GRE score by more than 10 points. They are not interested in knowing if their course leads to a significantly lower score.

6. (continued)

b) Since the P-value of 0.981 is high, fail to reject the null hypothesis. There is not enough evidence to conclude that the course improves scores for more than 50% of students by more than 10 points. Since the P-value is greater than 0.5, this sample actually had fewer than 50% of the students increase their score by more than 10 points.

c) Since the P-value of 0.019 is low, reject the null hypothesis. There is evidence that their course improves scores for more than 50% of students by more than 10 points.

Section 18.3

7. Hispanic origin.

a) H_0: The proportion of people in the county that are of Hispanic or Latino origin is 0.16. ($p = 0.16$)
H_A: The proportion of people in the county that are of Hispanic or Latino origin is different from 0.16. ($p \neq 0.16$)

b) **Randomization condition:** The 437 county residents were a random sample of all county residents.
10% condition: 437 is likely to be less than 10% of all county residents.
Success/Failure condition: $np_0 = (437)(0.16) = 69.92$ and $nq_0 = (437)(0.84) = 367.08$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we will model the sampling distribution of $\hat{p}$ with a Normal model and perform a one-proportion z-test.

c) $$\hat{p} = \frac{44}{437} = 0.101; \quad SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.16)(0.84)}{437}} \approx 0.018$$

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})} = \frac{0.101 - 0.16}{0.018} = -3.28; \quad \text{P-value} = 2 \cdot P(z < -3.28) = 0.001$$

d) Since the P-value = 0.001 is so low, reject the null hypothesis. There is evidence that the Hispanic/Latino population in this county differs from that as the nation as a whole. These data suggest that the proportion of Hispanic/Latino residents is, in fact, lower than the national proportion.

8. Empty houses.

a) H_0: The proportion of vacant houses in the county 0.114. ($p = 0.114$)
H_A: The proportion of vacant houses is different from 0.114. ($p \neq 0.114$)

b) **Randomization condition:** The 850 housing units were a random sample of all housing units in the county.
10% condition: 850 is likely less than 10% of all housing units in the county.
Success/Failure condition: $np_0 = (850)(0.114) = 96.9$ and $nq_0 = (850)(0.886) = 753.1$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we will model the sampling distribution of $\hat{p}$ with a Normal model and perform a one-proportion z-test.

c) $$\hat{p} = \frac{129}{850} = 0.152; \quad SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.114)(0.886)}{850}} \approx 0.011$$

$$z = \frac{\hat{p} - p_0}{SD(\hat{p})} = \frac{0.152 - 0.114}{0.011} = 3.45; \quad P\text{-value} = 2 \cdot P(z < 3.45) = 0.00056$$

8. (continued)

d) Since the P-value = 0.00056 is so low, reject the null hypothesis. There is evidence that the proportion of houses that are vacant in this county is different than the national proportion. These data suggest that the proportion of vacant houses is, in fact, higher than the national proportion.

Section 18.4

9. GRE performance again.

a) An increase in the mean score would mean that the mean difference (After − Before) is positive.

H_0: The mean of the differences in GRE scores (After − Before) is zero. $\left(\mu_{diff} = 0\right)$

H_A: The mean of the differences in GRE scores (After − Before) is greater than zero. $\left(\mu_{diff} > 0\right)$

b) Since the P-value, 0.65, is high, there is not convincing evidence that the mean difference in GRE scores (After − Before) is greater than zero. There is no evidence that the course is effective.

c) Because the P-value is greater than 0.5 and the alternative is one-sided (>0), we can conclude that the actual mean difference was less than 0. (If it had been positive, the probability to the right of that value would have to be less than 0.5.) This means that, in the sample of customers, the scores generally decreased after the course.

10. Marriage.

a) H_0: The mean age at which American men first marry is 23.3 years. $\left(\mu = 23.3\right)$

H_A: The mean age is greater than 23.3 years. $\left(\mu > 23.3\right)$

b) **Randomization condition:** The 40 men were selected randomly.
Nearly Normal condition: The population of ages of men at first marriage is likely to be skewed to the right. It is much more likely that there are men who marry for the first time at an older age than at an age that is very young. We should examine the distribution of the sample to check for serious skewness and outliers, but with a large sample of 40 men, it should be safe to proceed.

c) Since the conditions for inference are satisfied, we can model the sampling distribution of the mean age of men at first marriage with $N\left(23.3, \frac{\sigma}{\sqrt{n}}\right)$. Since we do not know σ, the standard deviation of the population, $\sigma\left(\bar{y}\right)$ will be estimated by $SE\left(\bar{y}\right) = \frac{s}{\sqrt{n}}$, and we will use a Student's *t*-model, with 40 – 1 = 39 degrees of freedom, $t_{39}\left(23.3, \frac{5.3}{\sqrt{40}}\right)$.

d) The mean age at first marriage in the sample was 24.2 years, with a standard deviation in age of 5.3 years. Use a one-sample *t*-test, modeling the sampling distribution of $\bar{y}$ with $t_{39}\left(23.3, \frac{5.3}{\sqrt{40}}\right)$.

$$t = \frac{\bar{y} - \mu_0}{SE\left(\bar{y}\right)} = \frac{24.2 - 23.3}{\frac{5.3}{\sqrt{40}}} \approx 1.07;\ \text{The P-value is 0.1447.}$$

e) If the mean age at first marriage is still 23.3 years, there is a 14.5% chance of getting a sample mean of 24.2 years or older simply from natural sampling variation.

f) Since the P-value = 0.1447 is high, we fail to reject the null hypothesis. There is no evidence to suggest that the mean age of men at first marriage has changed from 23.3 years, the mean in 1960.

11. Pizza.

If in fact the mean cholesterol of pizza eaters does not indicate a health risk, then only 7 out of every 100 samples would be expected to have mean cholesterol as high or higher than the mean cholesterol observed in the sample.

12. Golf balls.

If in fact this ball meets the velocity standard, then 34% of all samples tested would be expected to have mean speeds at least as high as the mean speed recorded in the sample.

Section 18.5

13. Bad medicine.

a) The drug may not be approved for use. People would miss out on a beneficial product and the company would miss out on potential sales.

b) The drug will go into production and people will suffer the side effect.

14. Expensive medicine.

a) People will pay more money for a drug that is no better than the old drug.

b) The new drug would not be used, so people would miss an opportunity for a more effective drug.

Chapter Exercises

15. Hypotheses.

a) H_0: The governor's "negatives" are 30%. ($p = 0.30$)
H_A: The governor's "negatives" are less than 30%. ($p < 0.30$)

b) H_0: The proportion of heads is 50%. ($p = 0.50$)
H_A: The proportion of heads is not 50%. ($p \neq 0.50$)

c) H_0: The proportion of people who quit smoking is 20%. ($p = 0.20$)
H_A: The proportion of people who quit smoking is greater than 20%. ($p > 0.20$)

16. More hypotheses.

a) H_0: The proportion of high school graduates is 40%. ($p = 0.40$)
H_A: The proportion of high school graduates is not 40%. ($p \neq 0.40$)

b) H_0: The proportion of cars needing transmission repair is 20%. ($p = 0.20$)
H_A: The proportion of cars is less than 20%. ($p < 0.20$)

c) H_0: The proportion of people who like the flavor is 60%. ($p = 0.60$)
H_A: The proportion of people who like the flavor is greater than 60%. ($p > 0.60$)

17. Negatives.

Statement d is the correct interpretation of a P-value. It talks about the probability of seeing the data, not the probability of the hypotheses.

18. Dice.

Statement d is the correct interpretation of a P-value. It correctly leads to rejecting the null hypothesis and talks about the probability of seeing the data, not the probability of the hypotheses.

19. Relief.

It is *not* reasonable to conclude that the new formula and the old one are equally effective. Furthermore, our inability to make that conclusion has nothing to do with the P-value. We cannot prove the null hypothesis (that the new formula and the old formula are equally effective), but can only fail to find evidence that would cause us to reject it. All we can say about this P-value is that there is a 27% chance of seeing the observed effectiveness from natural sampling variation if the new formula and the old one are equally effective.

20. Cars.

It is reasonable to conclude that a greater proportion of high schoolers have cars. If the proportion were no higher than it was a decade ago, there is only a 1.7% chance of seeing such a high sample proportion just from natural sampling variability.

21. He cheats?

a) Two losses in a row aren't convincing. There is a 25% chance of losing twice in a row, and that is not unusual.

b) If the process is fair, three losses in a row can be expected to happen about 12.5% of the time. $(0.5)(0.5)(0.5) = 0.125$.

c) Three losses in a row is still not a convincing occurrence. We'd expect that to happen about once every eight times we tossed a coin three times.

d) Answers may vary. Maybe 5 times would be convincing. The chances of 5 losses in a row are only 1 in 32, which seems unusual.

22. Candy.

a) $P(\text{first three vanilla}) = \left(\frac{6}{12}\right)\left(\frac{5}{11}\right)\left(\frac{4}{10}\right) \approx 0.091$

b) It seems reasonable to think there really may have been six of each. We would expect to get three vanillas in a row about 9% of the time. That's unusual, but not *that* unusual.

c) If the fourth candy was also vanilla, we'd probably start to think that the mix of candies was not 6 vanilla and 6 peanut butter. The probability of 4 vanilla candies in a row is:

$$P(\text{first four vanilla}) = \left(\frac{6}{12}\right)\left(\frac{5}{11}\right)\left(\frac{4}{10}\right)\left(\frac{3}{9}\right) \approx 0.03$$

We would only expect to get four vanillas in a row about 3% of the time. That's unusual.

23. Smartphones.

1) Null and alternative hypotheses should involve p, not $\hat{p}$.

2) The question is about *failing* to meet the goal. H_A should be $p < 0.96$.

3) The student failed to check $nq_0 = (200)(0.04) = 8$. Since $nq_0 < 10$, the Success/Failure condition is violated. Similarly, the 10% Condition is not verified.

4) $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.96)(0.04)}{200}} \approx 0.014$. The student used $\hat{p}$ and $\hat{q}$.

5) Value of z is incorrect. The correct value is $z = \frac{0.94 - 0.96}{0.014} \approx -1.43$.

6) P-value is incorrect. $P = P(z < -1.43) = 0.076$

7) For the P-value given, an incorrect conclusion is drawn. A P-value of 0.12 provides no evidence that the new system has failed to meet the goal. The correct conclusion for the corrected P-value is: Since the P-value of 0.076 is fairly low, there is weak evidence that the new system has failed to meet the goal.

24. Obesity 2016.

1) Null and alternative hypotheses should involve p, not $\hat{p}$.

2) The question asks if there is evidence that the 36.5% figure is *not accurate*, so a two-sided alternative hypothesis should be used. H_A should be $p \neq 0.365$.

3) The conditions are SRS, 750 < 10% of county population, $np_0 = (750)(0.365) = 273.75 \geq 10$, $nq_0 = (750)(0.635) = 476.25 \geq 10$

4) $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.365)(0.635)}{750}} \approx 0.018.$ Values of $\hat{p}$ and $\hat{q}$ were used.

5) Value of z is incorrect. The correct value is $z = \frac{0.304 - 0.365}{0.018} \approx -3.39.$

6) The correct, two-tailed P-value is $P = 2P(z < -3.39) = 0.0007$.

7) The P-value is misinterpreted. Since the P-value is so low, there is strong evidence that the proportion of adults in the county who are obese is different than the claimed 36.5%. In fact, our sample suggests that the proportion may be lower. There is only a 0.07% chance of observing a $\hat{p}$ as far from 0.365 as this simply from natural sampling variation.

25. Dowsing.

a) H_0: The percentage of successful wells drilled by the dowser is 30%. ($p = 0.30$)
H_A: The percentage of successful wells drilled is greater than 30%. ($p > 0.30$)

b) **Independence assumption:** There is no reason to think that finding water in one well will affect the probability that water is found in another, unless the wells are close enough to be fed by the same underground water source.
Randomization condition: This sample is not random, so hopefully the customers you check with are representative of all of the dowser's customers.
10% condition: The 80 customers sampled may be considered less than 10% of all possible customers.
Success/Failure condition: $np_0 = (80)(0.30) = 24$ and $nq_0 = (80)(0.70) = 56$ are both greater than 10, so the sample is large enough.

c) The sample of customers may not be representative of all customers, so we will proceed cautiously. A Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.30$ and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.30)(0.70)}{80}} \approx 0.0512.$$

The observed proportion of successful wells is $\hat{p} = \frac{27}{80} = 0.3375.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.3375 - 0.30}{\sqrt{\frac{(0.30)(0.70)}{80}}} \approx 0.73.$ P-value is 0.232.

d) If his dowsing has the same success rate as standard drilling methods, there is more than a 23% chance of seeing results as good as those of the dowser, or better, by natural sampling variation.

e) With a high P-value of 0.232, we fail to reject the null hypothesis. There is no evidence to suggest that the dowser has a success rate any higher than 30%.

26. Abnormalities.

a) H_0: The percentage of children with genetic abnormalities is 5%.($p = 0.05$)
H_A: The percentage of with genetic abnormalities is greater than 5%. ($p > 0.05$)

b) **Randomization condition:** This sample may not be random, but genetic abnormalities are plausibly independent. The sample is probably representative of all children, with regards to genetic abnormalities.
10% condition: The sample of 384 children is less than 10% of all children.
Success/Failure condition: $np_0 = (384)(0.05) = 19.2$ and $nq_0 = (384)(0.95) = 364.8$ are both greater than 10, so the sample is large enough.

c) The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.05$ and $SD(\hat{p}) = \sqrt{\frac{p_0q_0}{n}} = \sqrt{\frac{(0.05)(0.95)}{384}} \approx 0.0111$.

The observed proportion of children with genetic abnormalities is $\hat{p} = \frac{46}{384} \approx 0.1198$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{0.1198 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{384}}} \approx 6.28$.

The value of z is approximately 6.28, meaning that the observed proportion of children with genetic abnormalities is over 6 standard deviations above the hypothesized proportion. The P-value associated with this z score is 2×10^{-10}, essentially 0.

d) If 5% of children have genetic abnormalities, the chance of observing 46 children with genetic abnormalities in a random sample of 384 children is essentially 0.

e) With a P-value of this low, we reject the null hypothesis. There is strong evidence that more than 5% of children have genetic abnormalities.

f) We don't know that environmental chemicals cause genetic abnormalities. We merely have evidence that suggests that a greater percentage of children are diagnosed with genetic abnormalities now, compared to the 1980s.

27. Absentees.

a) H_0: The percentage of students in 2000 with perfect attendance the previous month is 34%. ($p = 0.34$)
H_A: The percentage of students in 2000 with perfect attendance the previous month is different from 34%. ($p \neq 0.34$)

b) **Randomization condition:** Although not specifically stated, we can assume that the National Center for Educational Statistics used random sampling.
10% condition: The 8302 students are less than 10% of all students.
Success/Failure condition: $np_0 = (8302)(0.34) = 2822.68$ and $nq_0 = (8302)(0.66) = 5479.32$ are both greater than 10, so the sample is large enough.

c) Since the conditions for inference are met, a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.34$ and $SD(\hat{p}) = \sqrt{\frac{p_0q_0}{n}} = \sqrt{\frac{(0.34)(0.66)}{8302}} \approx 0.0052$.

The observed proportion of perfect attendees is $\hat{p} = 0.33$.

We can perform a two-tailed one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{0.33 - 0.34}{\sqrt{\frac{(0.34)(0.66)}{8302}}} \approx -1.923$. P-value is 0.054. (0.055 using tables.)

27. (continued)

d) With a P-value of 0.0544, we reject the null hypothesis. There is some evidence to suggest that the percentage of students with perfect attendance in the previous month has changed in 2000.

e) This result is not meaningful. A difference this small, although statistically significant, is of little practical significance.

28. Educated mothers.

a) H_0: The percentage of students in 2000 whose mothers had graduated college is 31%. ($p = 0.31$)
H_A: The percentage of students is different than 31%. ($p \neq 0.31$)

b) **Randomization condition:** Although not specifically stated, we can assume that the National Center for Educational Statistics used random sampling.
10% condition: The 8368 students are less than 10% of all students.
Success/Failure condition: $np_0 = (8368)(0.31) = 2594.08$ and $nq_0 = (8368)(0.69) = 5773.92$ are both greater than 10, so the sample is large enough.

c) Since the conditions for inference are met, a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.31$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.31)(0.69)}{8368}} \approx 0.0051$.

The observed proportion of students whose mothers are college graduates is $\hat{p} = 0.32$.

We can perform a one-proportion two-tailed z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.32 - 0.31}{\sqrt{\frac{(0.31)(0.69)}{8368}}} \approx 1.978$. P-value is 0.048.

d) With a P-value of 0.048, we reject the null hypothesis. There is evidence to suggest that the percentage of students whose mothers are college graduates has changed since 1996. In fact, the evidence suggests that the percentage has increased.

e) This result is not meaningful. A difference this small, although statistically significant, is of little practical significance.

29. Contributions, please II.

a) H_0: The contribution rate is 5%. ($p = 0.05$)
H_A: The contribution rate is less than 5%. ($p < 0.05$)

b) **Randomization condition:** Potential donors were randomly selected.
10% condition: We will assume the entire mailing list has over 1,000,000 names.
Success/Failure condition: $np_0 = 5000$ and $nq_0 = 95{,}000$ 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 $p_0 = 0.05$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.05)(0.95)}{100{,}000}} \approx 0.0007$.

The observed contribution rate is $\hat{p} = \frac{4781}{100{,}000} = 0.04781$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.04781 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{100{,}000}}} \approx -3.178$. P-value is 0.00074.

c) Since the P-value = 0.00074 is low, we reject the null hypothesis. There is strong evidence that contribution rate for all potential donors is lower than 5%.

30. Take the offer II.

a) H_0: The success rate is 2%. ($p = 0.02$)
H_A: The success rate is something other than 2%. ($p \neq 0.02$)

b) **Randomization condition:** The sample was 50,000 randomly selected cardholders.
10% condition: We will assume that the number of cardholders is more than 500,000.
Success/Failure condition: $np_0 = 1000$ and $nq_0 = 49{,}000$ 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 $p_0 = 0.02$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.02)(0.98)}{50{,}000}} \approx 0.0006$.

The observed success rate is $\hat{p} = \frac{1184}{50{,}000} = 0.02368$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.02368 - 0.02}{\sqrt{\frac{(0.02)(0.98)}{50{,}000}}} \approx 5.878$. P-value < 0.0001.

c) Since $z = 5.878$, and the P-value is less than 0.0001, we reject the null hypothesis. There is strong evidence that success rate for all cardholders is not 2%. In fact, this sample suggests that the success rate is higher than 2%.

31. Pollution.

H_0: The percentage of cars with faulty emissions is 20%. ($p = 0.20$)
H_A: The percentage of cars with faulty emissions is greater than 20%. ($p > 0.20$)

Two conditions are not satisfied. 22 is greater than 10% of the population of 150 cars, and $np_0 = (22)(0.20) = 4.4$, which is not greater than 10. It's not advisable to proceed with a test.

32. Scratch and dent.

H_0: The percentage of damaged machines is 2%, and the warehouse is meeting the company goal. ($p = 0.02$)
H_A: The percentage of damaged machines is greater than 2%, and the warehouse is failing to meet the company goal. ($p > 0.02$)

An important condition is not satisfied. $np_0 = (60)(0.02) = 1.2$, which is not greater than 10. The Normal model is not appropriate for modeling the sampling distribution.

33. Twins.

H_0: The percentage of twin births to teenage girls is 3%. ($p = 0.03$)
H_A: The percentage of twin births to teenage girls differs from 3%. ($p \neq 0.03$)

Independence assumption: One mother having twins will not affect another. Observations are plausibly independent.
Randomization condition: This sample may not be random, but it is reasonable to think that this hospital has a representative sample of teen mothers, with regards to twin births.
10% condition: The sample of 469 teenage mothers is less than 10% of all such mothers.
Success/Failure condition: $np_0 = (469)(0.03) = 14.07$ and $nq_0 = (469)(0.97) = 454.93$ 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 $p_0 = 0.03$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.03)(0.97)}{469}} \approx 0.0079$.

33. (continued)

The observed proportion of twin births to teenage mothers is $\hat{p} = \dfrac{7}{469} \approx 0.015.$

We can perform a one-proportion z-test: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{0.015 - 0.03}{\sqrt{\dfrac{(0.03)(0.97)}{469}}} \approx -1.91.$

Since the P-value = 0.0556 is fairly low, we reject the null hypothesis. There is some evidence that the proportion of twin births for teenage mothers at this large city hospital is lower than the proportion of twin births for all mothers.

34. Football 2016.

H_0: The percentage of home team wins is 50%. ($p = 0.50$)
H_A: The percentage of home team wins is greater than 50%. ($p > 0.50$)

Independence assumption: Results of one game should not affect others.
Randomization condition: This season should be representative of other seasons, with regards to home team wins.
10% condition: 238 games represent less than 10% of all games, in all seasons.
Success/Failure condition: $np_0 = (238)(0.50) = 119$ and $nq_0 = (238)(0.50) = 119$ 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 $p_0 = 0.50$ and $SD(\hat{p}) = \sqrt{\dfrac{p_0 q_0}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{238}} \approx 0.03241.$

The observed proportion of home team wins is $\hat{p} = \dfrac{137}{238} = 0.5756$.

We can perform a one-proportion z-test: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{0.5756 - 0.5}{\sqrt{\dfrac{(0.5)(0.5)}{238}}} \approx 2.333.$

Since the P-value = 0.00981 is low, we reject the null hypothesis. There is strong evidence that the proportion of home team wins is greater than 50%. This provides evidence of a home team advantage.

35. WebZine.

H_0: The percentage of readers interested in an online edition is 25%. ($p = 0.25$)
H_A: The percentage of readers interested is greater than 25%. ($p > 0.25$)

Randomization condition: The magazine conducted an SRS of 500 current readers.
10% condition: 500 readers are less than 10% of all potential subscribers.
Success/Failure condition: $np_0 = (500)(0.25) = 125$ and $nq_0 = (500)(0.75) = 375$ 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 $p_0 = 0.25$ and $SD(\hat{p}) = \sqrt{\dfrac{p_0 q_0}{n}} = \sqrt{\dfrac{(0.25)(0.75)}{500}} \approx 0.0194.$

35. (continued)

The observed proportion of interested readers is $\hat{p} = \frac{137}{500} = 0.274$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.274 - 0.25}{\sqrt{\frac{(0.25)(0.75)}{500}}} \approx 1.24$.

Since the P-value = 0.1076 is high, we fail to reject the null hypothesis. There is little evidence to suggest that the proportion of interested readers is greater than 25%. The magazine should not publish the online edition.

36. Seeds.

H_0: The germination rate of the green bean seeds is 92%. ($p = 0.92$)
H_A: The germination rate of the green bean seeds is less than 92%. ($p < 0.92$)

Independence assumption: Seeds in a single packet may not germinate independently. They have been treated identically with regards to moisture exposure, temperature, etc. They may have higher or lower germination rates than seeds in general.
Randomization condition: The cluster sample of one bag of seeds was not random.
10% condition: 200 seeds is less than 10% of all seeds.
Success/Failure condition: $np_0 = (200)(0.92) = 184$ and $nq_0 = (200)(0.08) = 16$ are both greater than 10, so the sample is large enough.

The conditions have *not* been satisfied. We will assume that the seeds in the bag are representative of all seeds, and cautiously use a Normal model to model the sampling distribution of the proportion, with $p_0 = 0.92$ and

$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.92)(0.08)}{200}} \approx 0.0192$.

The observed proportion of germinated seeds is $\hat{p} = \frac{171}{200} = 0.855$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.855 - 0.92}{\sqrt{\frac{(0.92)(0.08)}{200}}} \approx -3.39$.

Since the P-value = 0.0004 is very low, we reject the null hypothesis. There is strong evidence that the germination rate of the seeds in less than 92%. We should use extreme caution in generalizing these results to all seeds, but the manager should be safe, and avoid selling faulty seeds. The seeds should be thrown out.

37. Women executives.

H_0: The proportion of female executives is similar to the overall proportion of female employees at the company. ($p = 0.40$)
H_A: The proportion of female executives is lower than the overall proportion of female employees at the company. ($p < 0.40$)

Independence assumption: It is reasonable to think that executives at this company were chosen independently.
Randomization condition: The executives were not chosen randomly, but it is reasonable to think of these executives as representative of all potential executives over many years.
10% condition: 43 executives are less than 10% of all executives at the company.
Success/Failure condition: $np_0 = (43)(0.40) = 17.2$ and $nq_0 = (43)(0.60) = 25.8$ are both greater than 10, so the sample is large enough.

37. (continued)

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $p_0 = 0.40$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.40)(0.60)}{43}} \approx 0.0747$.

The observed proportion is $\hat{p} = \frac{13}{43} \approx 0.302$.

Perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.302 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{43}}} \approx -1.31$.

Since the P-value = 0.0955 is high, we fail to reject the null hypothesis. There is little evidence to suggest proportion of female executives is any different from the overall proportion of 40% female employees at the company.

38. Jury.

H_0: The proportion of Hispanics called for jury duty is similar to the proportion of Hispanics in the county, 19%. ($p = 0.19$)
H_A: The proportion of Hispanics called for jury duty is less than the proportion of Hispanics in the county, 19%. ($p < 0.19$)

Independence assumption /Randomization condition: Assume that potential jurors were called randomly from all of the residents in the county. This is really what we are testing. If we reject the null hypothesis, we will have evidence that jurors are not called randomly.
10% condition: 72 people are less than 10% of all potential jurors in the county.
Success/Failure condition: $np_0 = (72)(0.19) = 13.68$ and $nq_0 = (72)(0.81) = 58.32$ 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 $p_0 = 0.19$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.19)(0.81)}{72}} \approx 0.0462$.

The observed proportion of Hispanics called for jury duty is $\hat{p} = \frac{9}{72} = 0.125$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.125 - 0.19}{\sqrt{\frac{(0.19)(0.81)}{72}}} \approx -1.41$.

Since the P-value = 0.0793 is somewhat high, we fail to reject the null hypothesis. We are not convinced that Hispanics are underrepresented in the jury selection system. However, this P-value isn't extremely high. There is some evidence that the selection process may be biased. We should examine some other groups called for jury duty and take a closer look.

39. Dropouts 2014.

H_0: The proportion of dropouts at this high school is similar to 6.5%, the proportion of dropouts nationally. ($p = 0.065$)
H_A: The proportion of dropouts at this high school is greater than 6.5%, the proportion of dropouts nationally. ($p > 0.065$)

Independence assumption /Randomization condition: Assume that the students at this high school are representative of all students nationally. This is really what we are testing. The dropout rate at this high school has traditionally been close to the national rate. If we reject the null hypothesis, we will have evidence that the dropout rate at this high school is no longer close to the national rate.

39. (continued)

10% condition: 1782 students are less than 10% of all students nationally.
Success/Failure condition: $np_0 = (1782)(0.065) = 115.83$ and $nq_0 = (1782)(0.935) = 1666.17$ 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, $p_0 = 0.065$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.065)(0.935)}{1782}} \approx 0.00584.$

The observed proportion of dropouts is $\hat{p} = \frac{130}{1782} \approx 0.0729517.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.0729517 - 0.065}{\sqrt{\frac{(0.065)(0.935)}{1782}}} \approx 1.362.$

Since the P-value = 0.0866 is not low, we fail to reject the null hypothesis. There is little evidence that the dropout rate at this high school is significantly higher than 6.5%.

40. Acid rain.

H_0: The proportion of trees with acid rain damage in Hopkins Forest is 15%, the proportion of trees with acid rain damage in the Northeast. ($p = 0.15$)
H_A: The proportion of trees with acid rain damage in Hopkins Forest is greater than 15%, the proportion of trees with acid rain damage in the Northeast. ($p > 0.15$)

Independence assumption /Randomization condition: Assume that the trees in Hopkins Forest are representative of all trees in the Northeast. This is really what we are testing. If we reject the null hypothesis, we will have evidence that the proportion of trees with acid rain damage is greater in Hopkins Forest than the proportion in the Northeast.
10% condition: 100 trees are less than 10% of all trees.
Success/Failure condition: $np_0 = (100)(0.15) = 15$ and $nq_0 = (100)(0.85) = 85$ 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 $p_0 = 0.15$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.15)(0.85)}{100}} \approx 0.0357.$

The observed proportion of damaged trees is $\hat{p} = \frac{25}{100} = 0.25.$

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.25 - 0.150}{\sqrt{\frac{(0.150)(0.850)}{100}}} \approx 2.80.$

Since the P-value = 0.0026 is low, we reject the null hypothesis. There is strong evidence that the trees in Hopkins forest have a greater proportion of acid rain damage than the 15% reported for the Northeast.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{25}{100}\right) \pm 1.96 \sqrt{\frac{\left(\frac{25}{100}\right)\left(\frac{75}{100}\right)}{100}} = (0.165, 0.335)$$

We are 95% confident that the interval 0.165 to 0.335 captures the true proportion of trees in the Hopkins forest that are damaged by acid rain, which is higher than the 15% reported for the Northeast.

41. Lost luggage.

H_0: The proportion of lost luggage returned the next day is 90%. ($p = 0.90$)
H_A: The proportion of lost luggage returned is lower than 90%. ($p < 0.90$)

Independence assumption: It is reasonable to think that the people surveyed were independent with regards to their luggage woes.
Randomization condition: Although not stated, we will hope that the survey was conducted randomly, or at least that these air travelers are representative of all air travelers for that airline.

10% condition: 122 air travelers are less than 10% of all air travelers on the airline.
Success/Failure condition: $np_0 = (122)(0.90) = 109.8$ and $nq_0 = (122)(0.10) = 12.2$ 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 $p_0 = 0.90$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.90)(0.10)}{122}} \approx 0.0272$.

The observed proportion of lost luggage is $\hat{p} = \frac{103}{122} \approx 0.844$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.844 - 0.90}{\sqrt{\frac{(0.90)(0.10)}{122}}} \approx -2.05$.

Since the P-value = 0.0201 is low, we reject the null hypothesis. There is evidence that the proportion of lost luggage returned the next day is lower than the 90% claimed by the airline.

42. TV ads.

H_0: The proportion of respondents who recognize the name is 40%.($p = 0.40$)
H_A: The proportion is more than 40%. ($p > 0.40$)

Randomization condition: The pollster contacted the 420 adults randomly.
10% condition: A sample of 420 adults is less than 10% of all adults.
Success/Failure condition: $np_0 = (420)(0.40) = 168$ and $nq_0 = (420)(0.60) = 252$ 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 $p_0 = 0.40$ and $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{(0.40)(0.60)}{420}} \approx 0.0239$.

The observed proportion is $\hat{p} = \frac{181}{420} \approx 0.431$.

We can perform a one-proportion z-test: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.431 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{420}}} \approx 1.29$.

Since the P-value = 0.0977 is fairly high, we fail to reject the null hypothesis. There is little evidence that more than 40% of the public recognizes the product. Don't run commercials during the Super Bowl!

43. John Wayne.

a) H_0: The death rate from cancer for people working on the film was similar to that predicted by cancer experts, 30 out of 220.
H_A: The death rate from cancer for people working on the film was higher than the rate predicted by cancer experts.

The conditions for inference are not met, since this is not a random sample. We will assume that the cancer rates for people working on the film are similar to those predicted by the cancer experts, and a Normal model can be used to model the sampling distribution of the rate, with $p_0 =$ 30/220 and

$$SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{\left(\frac{30}{220}\right)\left(\frac{190}{220}\right)}{220}} \approx 0.0231.$$

The observed cancer rate is $\hat{p} = \dfrac{46}{220} \approx 0.209.$

We can perform a one-proportion z-test: $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \dfrac{\frac{46}{220} - \frac{30}{220}}{\sqrt{\frac{\left(\frac{30}{220}\right)\left(\frac{190}{220}\right)}{220}}} = 3.14.$

Since the P-value = 0.0008 is very low, we reject the null hypothesis. There is strong evidence that the cancer rate is higher than expected among the workers on the film.

b) This does not prove that exposure to radiation may increase the risk of cancer. This group of people may be atypical for reasons that have nothing to do with the radiation.

44. AP Stats 2016.

a) Her students scored about 0.40 standard deviations above the national rate. This is not a large difference, less than half a standard deviation above the national rate. Even if her students scored no better than the national rate, we would expect many samples of size 54 to have rates this high or higher, just due to random chance.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{\frac{34}{54} - 0.603}{\sqrt{\frac{(0.603)(0.397)}{54}}} = 0.4000$$

b) Even without conducting a hypothesis test, we know that the P-value for this test would be quite high. The teacher has no reason to be overly pleased. Her students did have a higher rate of scores of 3 or higher, but not so high that the results could not be attributed to sampling variability.

45. Normal temperature again.

a) H_0: Mean body temperature is 98.6°F, as commonly assumed. ($\mu = 98.6$°F)
H_A: Mean body temperature is not 98.6°F. ($\mu \neq 98.6$°F)

b) **Randomization condition:** The adults were randomly selected.
Nearly Normal condition: The sample of 52 adults is large, and the histogram shows no serious skewness, outliers, or multiple modes.

The people in the sample had a mean temperature of 98.285° and a standard deviation in temperature of 0.6824°. Since the conditions are satisfied, the sampling distribution of the mean can be modeled by a Student's t-model, with 52 – 1 = 51 degrees of freedom, $t_{51}\left(98.6, \dfrac{0.6824}{\sqrt{52}}\right)$.

45. (continued)

We will use a one-sample t-test: $t = \dfrac{\bar{y} - \mu_0}{SE(\bar{y})} = \dfrac{98.285 - 98.6}{\frac{0.6824}{\sqrt{52}}} \approx -3.33.$

c) Since the P-value = 0.0016 is low, we reject the null hypothesis. There is strong evidence that the true mean body temperature of adults is not 98.6°F. This sample would suggest that it is significantly lower.

46. Hot dogs again.

a) H_0: Mean sodium content is 325 mg. $(\mu = 325)$
H_A: Mean sodium content is less than 325 mg. $(\mu < 325)$

The hot dogs in the sample have a mean sodium content of 322.0 mg and standard deviation in sodium content of 18 mg. We are assuming that the assumptions and conditions for inference are satisfied, so the sampling distribution of the mean can be modeled by a Student's t-model, with 40 – 1 = 39 degrees of freedom, $t_{39}\left(325, \dfrac{18}{\sqrt{40}}\right)$.

We will use a one-sample t-test: $t = \dfrac{\bar{y} - \mu_0}{SE(\bar{y})} = \dfrac{322.0 - 325}{\frac{18}{\sqrt{40}}} \approx -1.05$

Since the P-value = 0.149 is high, we fail to reject the null hypothesis. There is no convincing evidence that the true mean sodium content of hot dogs is significantly less than 325 mg.

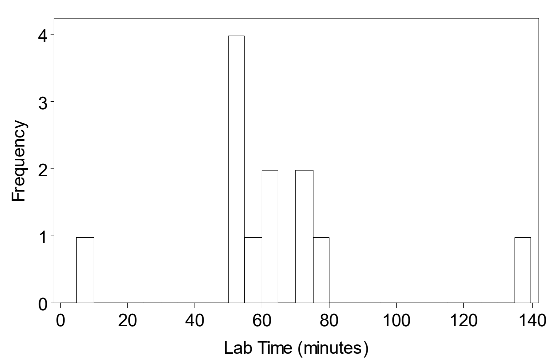
b) No, a larger sample will not ensure that the regulation is met. If their mean sodium content is actually less than the limit, a larger sample size will help properly detect that. However, we do not know for certain that the mean is lower. We only have the results of one sample.

47. More Pizza.

If the mean cholesterol level really does not exceed the value considered to indicate a health risk, there is a 7% probability that a random sample of this size would have a mean as high as (or higher than) the mean in this sample.

48. Computer lab fees again.

a) The histogram of the number of minutes spent in the computer lab is shown below. There appear to be two outliers: one time, 8 minutes, is much lower than the others, and another time, 136 minutes is much higher than the others. If the outliers are included, the conditions for inference are violated.



48. (continued)

b) H_0: The mean number of minutes spent in the computer lab is 55. ($\mu = 55$)

H_A: The mean number of minutes spent in the computer lab has increased, and is greater than 55. ($\mu > 55$)

Randomization condition: The 12 students were selected at random,
Nearly Normal condition: The histogram of the computer lab times in the sample is not nearly normal. There are two outliers.

The students in the sample spent a mean time of 63.25 minutes in the computer lab, with a standard deviation of 28.9266 minutes. The assumptions and conditions for inference are not satisfied, but we have been instructed to proceed both with and without the outliers. The sampling distribution of the mean can be modeled by a Student's t-model, with 12 – 1 = 11 degrees of freedom, $t_{11}\left(55, \frac{28.9266}{\sqrt{12}}\right)$.

We will use a one-sample t-test: $t = \dfrac{\bar{y} - \mu_0}{SE(\bar{y})} = \dfrac{63.25 - 55}{\frac{28.9266}{\sqrt{12}}} \approx 0.988.$

Since the P-value = 0.172 is high, we fail to reject the null hypothesis. There is no convincing evidence that the true mean time spent in the computer lab is significantly longer than 55 minutes.

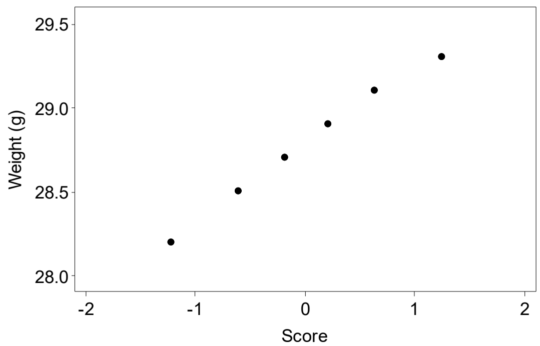
With the removal of the outliers, the assumptions and conditions for inference are satisfied, since the distribution of computer lab times is now reasonably unimodal and symmetric. The students in the sample spent a mean time of 61.5 minutes in the computer lab, with a standard deviation of 9.5946 minutes. The sampling distribution of the mean can be modeled by a Student's t-model, with 10 – 1 = 9 degrees of freedom, $t_9\left(55, \frac{9.5946}{\sqrt{10}}\right)$.

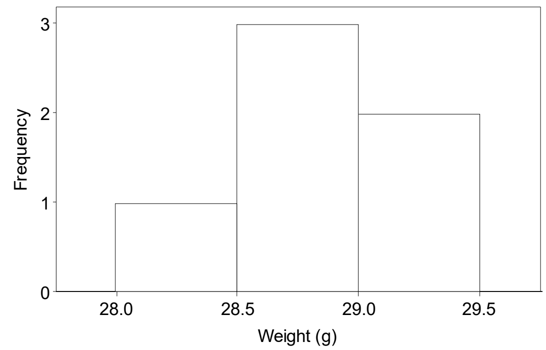
We will use a one-sample t-test: $t = \dfrac{\bar{y} - \mu_0}{SE(\bar{y})} = \dfrac{61.5 - 55}{\frac{9.5946}{\sqrt{10}}} \approx 2.142.$

Since the P-value = 0.03 is low, we reject the null hypothesis. There is convincing evidence that the true mean time spent in the computer lab is significantly longer than 55 minutes. (Notice that with the removal of the outliers, we have evidence of a mean time in the computer lab significantly longer than 55 minutes, even though that mean time, 61.5 minutes, is shorter than the mean time with the outliers, 63.5 minutes. The removal of the outliers results in much less variability in the computer lab times, as shown by the much smaller standard deviation of times).

49. More Ruffles.

a) **Randomization condition:** The 6 bags were not selected at random, but it is reasonable to think that these bags are representative of all bags of chips.
Nearly Normal condition: The Normal probability plot is reasonably straight, and the histogram of the weights of chips in the sample is nearly normal.





b) $\bar{y} \approx 28.78$ grams, $s \approx 0.40$ grams

c) H_0: The mean weight of Ruffles bags is 28.3 grams. ($\mu = 28.3$)
H_A: The mean weight of Ruffles bags is not 28.3 grams. ($\mu \neq 28.3$)

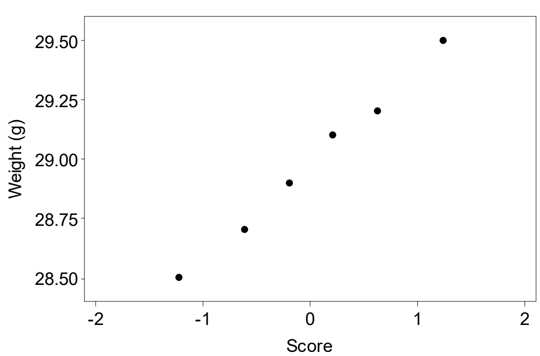
Since the conditions for inference have been satisfied, the sampling distribution of the mean can be modeled by a Student's t-model, with $6 - 1 = 5$ degrees of freedom, $t_5\left(28.3, \frac{0.40}{\sqrt{6}}\right)$.

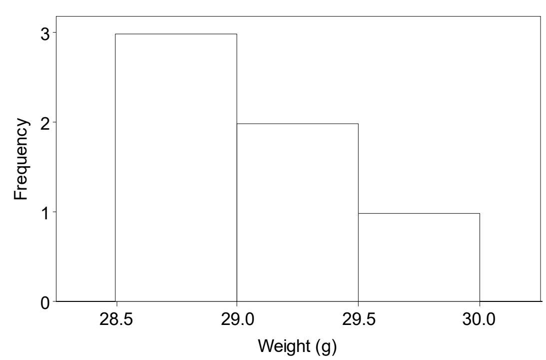
We will use a one-sample t-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{28.78 - 28.3}{\frac{0.40}{\sqrt{6}}} \approx 2.94.$

Since the P-value = 0.032 is low, we reject the null hypothesis. There is convincing evidence that the true mean weight of bags of Ruffles potato chips is not 28.3 grams. This sample suggests that the true mean weight is significantly higher.

50. More Doritos.

a) **Randomization condition:** The 6 bags were not selected at random, but it is reasonable to think that these bags are representative of all bags.
Nearly Normal condition: The Normal probability plot is reasonably straight. Although the histogram of the weights of chips in the sample is not symmetric, any apparent "skewness" is the result of a single bag of chips. It is safe to proceed.





50. (continued)

b) $\bar{y} \approx 28.98$ grams, $s \approx 0.36$ grams

c) H_0: The mean weight of Doritos bags is 28.3 grams. $\mu = 28.3$
H_A: The mean weight of Doritos bags is not 28.3 grams. $\mu \neq 28.3$

Since the conditions for inference have been satisfied, the sampling distribution of the mean can be modeled by a Student's *t*-model, with 6 – 1 = 5 degrees of freedom, $t_5\left(28.3, \frac{0.36}{\sqrt{6}}\right)$.

We will use a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{28.98 - 28.3}{\frac{0.36}{\sqrt{6}}} \approx 4.65.$

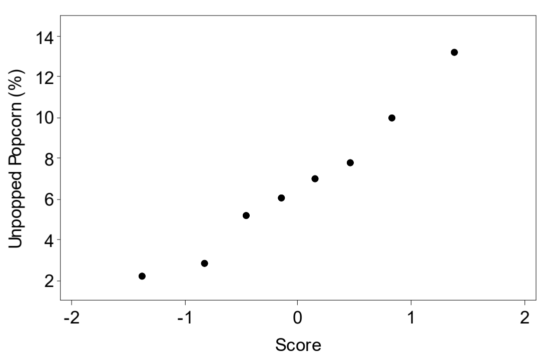
Since the P-value = 0.0056 is low, we reject the null hypothesis. There is convincing evidence that the true mean weight of bags of Doritos potato chips is not 28.3 grams. This sample suggests that the true mean weight is significantly higher.

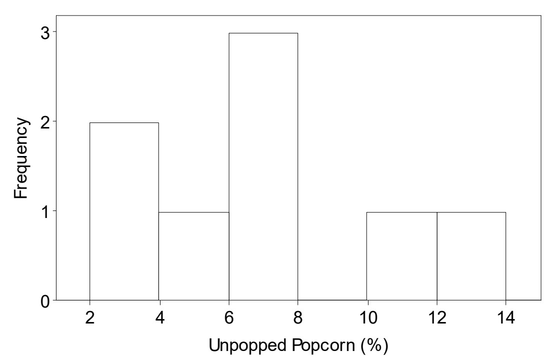
51. More Popcorn.

H_0: The mean proportion of unpopped kernels is 10%. $(\mu = 10)$

H_A: The mean proportion of unpopped kernels is lower than 10%. $(\mu < 10)$

Randomization condition: The 8 bags were randomly selected.
Nearly Normal condition: The histogram of the percentage of unpopped kernels is unimodal and roughly symmetric, and the Normal probability plot is reasonably straight.





The bags in the sample had a mean percentage of unpopped kernels of 6.775 percent and a standard deviation in percentage of unpopped kernels of 3.637 percent. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean percentage of unpopped kernels with a Student's *t*-model, with 8 – 1 = 7 degrees of freedom, $t_7\left(10, \frac{3.637}{\sqrt{8}}\right)$.

We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{6.775 - 10}{\frac{3.637}{\sqrt{8}}} \approx -2.51.$

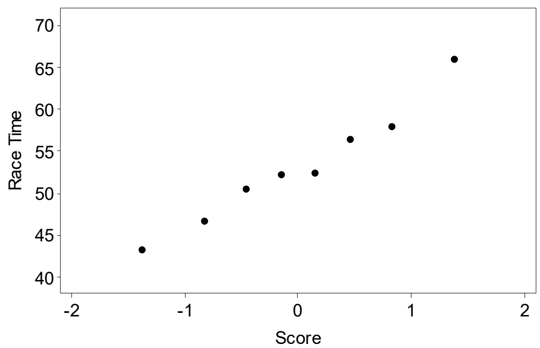
Since the P-value = 0.0203 is low, we reject the null hypothesis. There is evidence to suggest the mean percentage of unpopped kernels is less than 10% at this setting.

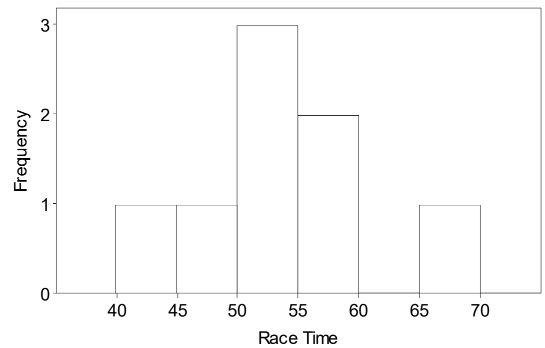
52. More ski wax.

H_0: The mean time was 55 seconds. $(\mu = 55)$

H_A: The mean time was less than 55 seconds. $(\mu < 55)$

Independence assumption: Since the times are not randomly selected, we will assume that the times are independent, and representative of all times.
Nearly Normal condition: The histogram of the times is unimodal and roughly symmetric, and the Normal probability plot is reasonably straight.





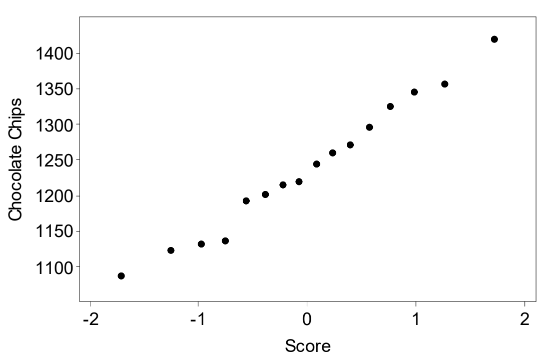
The times in the sample had a mean of 53.1 seconds and a standard deviation of 7.029 seconds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean time with a Student's *t*-model, with 8 – 1 = 7 degrees of freedom, $t_7\left(55, \frac{7.029}{\sqrt{8}}\right)$.

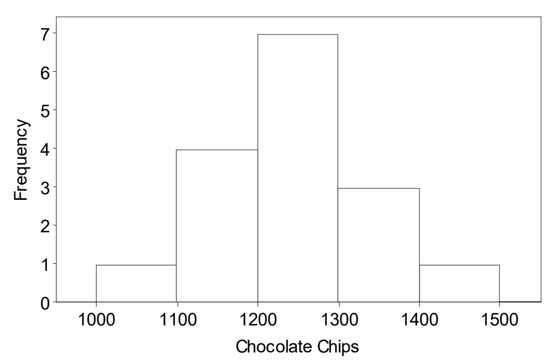
We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{53.1 - 55}{\frac{7.029}{\sqrt{8}}} \approx -0.7646$.

Since the P-value = 0.2347 is high, we fail to reject the null hypothesis. There is no evidence to suggest the mean time is less than 55 seconds. He should not buy the new ski wax.

53. Chips Ahoy! again.

a) Randomization condition: The bags of cookies were randomly selected.
Nearly Normal condition: The Normal probability plot is reasonably straight, and the histogram of the number of chips per bag is unimodal and symmetric.





53. (continued)

The bags in the sample had a mean number of chips of 1238.19, and a standard deviation of 94.282 chips. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean number of chips with a Student's t-model, with 16 – 1 = 15 degrees of freedom, $t_{15}\left(1000, \frac{94.282}{\sqrt{16}}\right)$.

We will perform a one-sample t-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{1238.19 - 1000}{\frac{94.282}{\sqrt{16}}} \approx 10.1.$

b) H₀: The mean number of chips per bag is 1000. $(\mu = 1000)$

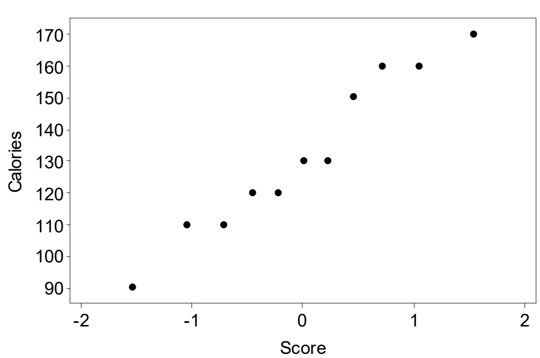
H_A: The mean number of chips per bag is greater than 1000. $(\mu > 1000)$

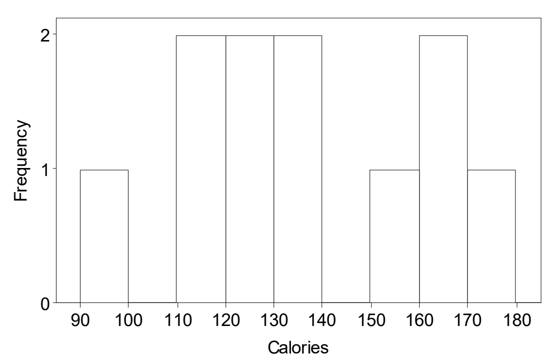
Since the P-value < 0.0001 is low, we reject the null hypothesis. There is convincing evidence that the true mean number of chips in bags of Chips Ahoy! Cookies is greater than 1000.

However, since the "1000 Chip Challenge" is about individual bags, not means, the claim made by Nabisco may not be true. They claim that *all* bags of cookies have at least 1000 chips. This test doesn't really answer that question. (If the mean was around 1188 chips, the low end of our confidence interval from the question in the last chapter, and the standard deviation of the population was about 94 chips, our best estimate obtained from our sample, a bag containing 1000 chips would be about 2 standard deviations below the mean. This is not likely to happen, but not an outrageous occurrence. These data do not provide evidence that the "1000 Chip Challenge" is true).

54. More yogurt.

a) **Randomization condition:** The brands of vanilla yogurt may not be a random sample, but they are probably representative of all brands of yogurt.
Independence assumption: The Randomization Condition is designed to check the reasonableness of the assumption of independence. We had some trouble verifying this condition. But is the calorie content per serving of one brand of yogurt likely to be associated with that of another brand? Probably not. We're okay.
Nearly Normal condition: The Normal probability plot is reasonably straight, and the histogram of the number of calories per serving is roughly unimodal (though somewhat uniform) and symmetric with no outliers.





54. (continued)

b) H_0: The mean number of calories is 120. $(\mu = 120)$

H_A: The mean number of calories is not 120. $(\mu \neq 120)$

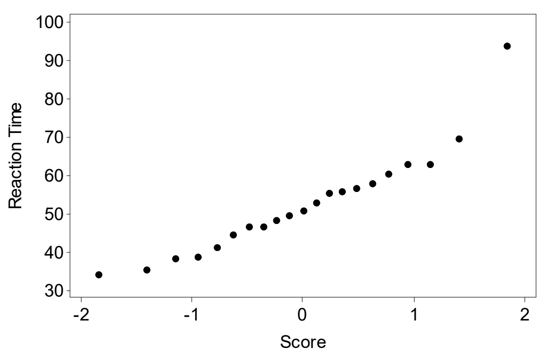
The brands in the sample had a mean calorie content of 131.82 calories, and a standard deviation of 25.23 calories. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean calorie content with a Student's *t*-model, with 11 – 1 = 10 degrees of freedom, $t_{10}\left(120, \frac{25.23}{\sqrt{11}}\right)$.

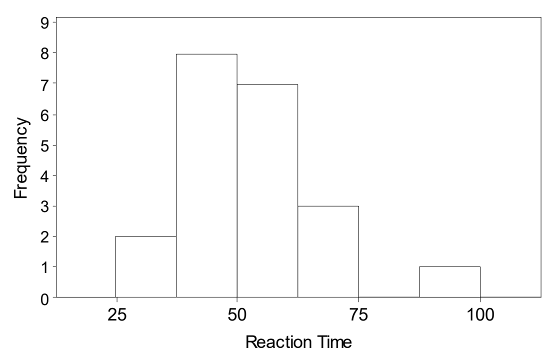
We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{131.82 - 120}{\frac{25.23}{\sqrt{11}}} \approx 1.55.$

Since the P-value = 0.15 is high, we fail to reject the null hypothesis. There is little evidence that the true mean number of calories in yogurt is different from 120 calories.

55. Maze.

a) **Independence assumption:** It is reasonable to think that the rats' times will be independent, as long as the times are for different rats.
Nearly Normal condition: There is an outlier in both the Normal probability plot and the histogram that should probably be eliminated before continuing the test. One rat took a long time to complete the maze.





b) H_0: The mean time for rats to complete this maze is 60 seconds. $(\mu = 60)$

H_A: The mean time for rats to complete this maze is at most 60 seconds. $(\mu < 60)$

The rats in the sample finished the maze with a mean time of 52.21 seconds and a standard deviation in times of 13.5646 seconds. We will model the sampling distribution of the mean time in which rats complete the maze with a Student's *t*-model, with 21 – 1 = 20 degrees of freedom, $t_{20}\left(60, \frac{13.5646}{\sqrt{21}}\right)$.

We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{52.21 - 60}{\frac{13.5646}{\sqrt{21}}} \approx -2.63.$

Since the P-value = 0.008 is low, we reject the null hypothesis. There is evidence that the mean time required for rats to finish the maze is less than 60 seconds.

55. (continued)

Without the outlier, the rats in the sample finished the maze with a mean time of 50.13 seconds and standard deviation in times of 9.90 seconds. Since the conditions for inference are satisfied, we can model the sampling distribution of the mean time in which rats complete the maze with a Student's *t*-model, with $20 - 1 = 19$ degrees of freedom, $t_{19}\left(60, \frac{9.90407}{\sqrt{20}}\right)$. We will use a one-sample *t*-test.

This test results in a value of $t = -4.46$, and a one-sided P-value = 0.0001. Since the P-value is low, we reject the null hypothesis. There is evidence that the mean time required for rats to finish the maze is less than 60 seconds.

According to both tests, there is evidence that the mean time required for rats to complete the maze is less than 60 seconds. This maze meets the requirement that the time be at most one minute, on average. The original question had a requirement that the maze take about one minute. If we performed a two-tailed test instead, the P-values would still be very small, and we would reject the null hypothesis that the maze completion time was one minute, on average. There is evidence that the maze does not meet the "one-minute average" requirement. It should be noted that the test without the outlier is the appropriate test. The one slow rat made the mean time required seem much higher than it probably was.

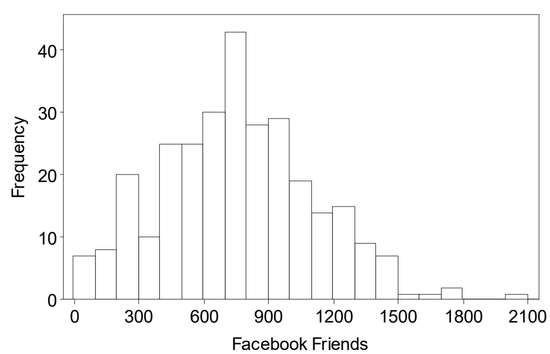
56. Facebook friends.

H_0: The mean number of Facebook friends at this school is 649. $(\mu = 649)$

H_A: The mean number of Facebook friends at this school is higher than 649. $(\mu > 649)$

Randomization condition: We don't know whether this sample was taken randomly.
Nearly Normal condition: The histogram of the number of Facebook friends is unimodal, and skewed to the right, but the sample size is large enough for the Central Limit Theorem to apply.

The students in the sample had a mean number of Facebook friends of 751.163 friends, and a standard deviation of 357.321 friends. As long as this turns out to be random sample of students at this high school, the conditions for inference are satisfied, and we can model the sampling distribution of the mean number of Facebook friends with a Student's *t*-model, with $294 - 1 = 293$ degrees of freedom, $t_{293}\left(649, \frac{357.321}{\sqrt{294}}\right)$.



We will perform a one-sample *t*-test: $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})} = \frac{751.136 - 649}{\frac{357.321}{\sqrt{294}}} \approx 4.90.$

Since the P-value < 0.0001 is low, we reject the null hypothesis. There is convincing evidence that the true mean number of Facebook friends at this high school is significantly higher than the average of 649 for all 18-24-year-olds.

57. Maze revisited.

a) To find the 99% bootstrap confidence interval, use the 0.5th percentile and the 99.5th percentile. We are 99% confident that the true mean maze completion time is between 45.271 and 59.939 seconds.

b) The sampling distribution is slightly skewed to the right because of the outlier at 93.8 seconds.

c) Because of the outlier, the confidence interval is not symmetric around the sample mean of 52.21 seconds.

d) Because 60 seconds is not in the interval, it is not plausible that the mean time is 60 seconds. The mean time appears to be less than 60 seconds.

e) We want the proportion of cases farther than 7.79 seconds from 60 seconds. On the left, that is fewer than 0.05% of cases. On the right, it is fewer than 0.5% of cases. We might estimate the P-value at about 0.004.

58. Facebook friends again.

a) The value 0 is far into a tail of this distribution so we know that P-value < 0.001, but because there are only 1000 bootstrap samples, we can't be more precise than that.

b) If this group of students can be considered a random sample of students at his school, we have convincing evidence that the mean number of Facebook friends for his school is not 649. It appears to be significantly higher.

59. Cholesterol.

a) The mean total cholesterol is 234.701 mg/dL

b) (Answers may vary slightly, depending on your bootstrap results.) We are 95% that the true mean total cholesterol level of adults is between 232.35 and 237.26 mg/dL.

c) (Answers may vary slightly, depending on your bootstrap results.) In our bootstrapped test, bootstrapped means as low as the sample mean of 234.701 never occurred. The bootstrapped P-value is < 0.001 using 1000 samples.

d) Both the interval and the test indicate that a mean of 240 mg/dL is not plausible. The mean cholesterol level of the Framingham participants is lower than 240 mg/dL by an amount too large to be attributed to random variability.

60. BMI.

a) The mean BMI for the men was 25.3.

b) (Answers may vary slightly, depending on your bootstrap results.) We are 95% confident that the mean BMI for American men is between 24.89 and 25.68.

c) (Answers may vary slightly, depending on your bootstrap results.) In our bootstrapped test, 55 of the 1000 bootstrapped means were farther away from 24.9 than the sample mean of 25.3. The bootstrapped P-value is 0.055.

d) The 95% interval just barely contains 24.9 and the P-value is just over 0.05. Both point to the fact that we have some, but not strong, evidence that the mean BMI is different from 24.9.