Introductory Statistics Explained (1.09)
Answers to Exercises

Contents

Chapter 1

Introduction

(No exercises.)

Chapter 2

Gathering Data

J.B.'s strongly suggested exercises: 1, 2, 3, 4, 5, 6, 8, 9, 12, 15, 17, 18, 21, 24, 29, 31, 32

2.1 Introduction

2.2 Populations and Samples, Parameters and Statistics

1. A parameter is a numerical characteristic of a *population*, whereas a statistic is a numerical characteristic of a *sample*. The value of a parameter is typically unknown, but the value of a statistic is known once the sample is taken.

2. (a) All three-year-old walleye in this lake.
   (b) The 52 fish caught in the netting program.
   (c) 273 is a statistic, since it is a value obtained from a *sample*.
   (d) The parameter of interest is the true mean weight of all three-year-old walleye in this lake.
   (e) No, not in a practical sense. The true mean weight is not a knowable quantity, as to find it we'd have to catch every three-year-old walleye in this lake.

2.3 Types of Sampling

2.3.1 Simple Random Sampling

3. (a) The 200 pacemakers in the shipment.

(b) The 8 pacemakers.
(c) 0% (a numerical characteristic of a sample)
(d) 10% (a numerical characteristic of the population).
(e) No. Only the top pacemakers were drawn (not all possible samples were equally possible).

4. (a) True. Each possible sample of size 10 had the same chance of being selected, and the sample obtained is thus a simple random sample.
   (b) True.
   (c) True.
   (d) True (since we are sampling 10 from a population of 50).

2.3.2 Other Types of Random Sampling

5. Statements a) and b) are true, and c) is false. Sample I is a simple random sample. Sample II is a stratified random sample. Both sampling designs are reasonable.

2.4 Experiments and Observational Studies

6. (a) It is an experiment.
   (b) The 20 volunteers.
   (c) The procedures (A and B).
   (d) Time until death.
   (e) Yes, since it is a randomized experiment.

7. (a) It is an experiment.
   (b) The types of corn.
   (c) The yield of the corn.
   (d) The plots of land.

8. No, it would not be reasonable. It's an observational study, and as such does not provide strong evidence of a causal link. There may appear to be an *association*, but this relationship could easily be caused by other variables. For example, perhaps uninvolved parents are more likely to have children that watch lots of TV, and more likely to have children that are arrested.

2.5 Chapter Exercises

9. (a) All students at this university.
   (b) No, it is not a simple random sample. Sampling methods like this are sometimes called *convenience* sampling.
   (c) The volunteers may be fundamentally different from other students at the university. (Perhaps the small giveaways and a chance of a $100 gift card lured them in because they had very little money.) These volunteers may tend to spend more or less on fast food than other students at the university. And since this was a survey, the respondents may lie, or may not remember how much they spend on fast food.
   (d) No, it would not be possible. We may be able to draw a simple random sample of students, but we cannot force these students to participate in the study.

10. (a) All lizards of these species in this region of Inner Mongolia.
    (b) No, this is not simple random sampling. Only lizards the researchers could catch were included in the sample.
    (c) The researchers would have an easier time catching certain types of lizard, and thus certain types of lizard would tend to be overrepresented in the sample. (Perhaps larger lizards are easier to catch, or smaller lizards are easier to catch.) It is hard to say how much of a problem this might be, but there are definitely possible biases in this study. But there is no better alternative—the researchers simply have to catch the lizards by hand.

11. (a) If patients knew they were receiving vitamin C rather than the placebo, they may be more confident that it is helping to reduce the duration and severity of colds. This may show up in their reporting of the results, even if there is no real effect. If the researchers knew which treatment the patient was receiving, they may be tempted (consciously or subconsciously) to tweak the results to reflect what they want to show.
    (b) This may or may not be a problem, depending on the reasons people had for dropping out of the study. The dropouts will not cause any bias if they are dropping out for reasons unrelated to the study. (For example, if they move out of town or die in an unrelated accident.) If they drop out for reasons related to the study then this may very well cause substantial bias. (For example, if they drop out because the vitamin C is making them too sick from colds to report the results.) In the original study, the researchers followed up with the dropouts, and found that most were dropping out because they were getting bored of the study and did not want to continue.

12. (a) The weight of the tumour is the response variable.
    (b) The exercise variable (forced exercise, no forced exercise) is the explanatory

variable.

(c) This will not cause any bias as long as the reason for the rat catching his foot on the treadmill is unrelated to the study variables. In the absence of further information, this is likely the case, so the results of the study are still valid. But if the reason for the rat catching his foot was related to the study variables, then this may result in substantial bias. (For example, if the rat caught his foot on the treadmill because of a lack of mobility due to a massive tumour, then omitting this rat would be problematic and may lead to misleading results.)

13. (a) Since only the numbers 1 through 7 count, the sample is made up of the numbers 7, 1, and 3. 98**7**97 **1**8**3**34. This corresponds to the countries Panama, Belize, El Salvador.

(b) In this case, yes. But only because none of these country names share the same first letter.

14. The alphabetical list (c) would provide the sample that has the least bias. (But this is open to debate.)

The sample in a) has many sources of potential bias, including:

- Only students that showed up early to the lecture following a midterm were included in the sample. Students that show up to the lecture after a midterm may very well differ from the rest of the class in a fundamental way.

- The lecture was on a Friday morning, and students attending Friday morning classes may differ from the rest of the class in a fundamental way.

- What students *say* their mark is on a test may very well differ from their mark on the test. (Students might have a tendency say they got a higher mark than they did.)

The sample in b) has many sources of potential bias, including:

- Only students in line for student loans are in the sample.

- The lecture was on a Friday morning, and students attending Friday morning classes may differ from the rest of the class in a fundamental way.

- What students *say* their parents' income is may be quite different from what their parents' income actually is.

While there are possible problems with sample c), they are likely not as severe as for the other two sampling designs. There is possibly a relationship between the first letter of the last name and the proportion of students that are male. For example, certain cultures may send males to university more often than females, and have last names that tend to start with a letter near the beginning of the alphabet. But these issues are likely not as severe as for the first 2 sampling designs.

15. (a) All full-time female graduate students at the university.
    (b) The 67 students that returned the survey.
    (c) The proportion of all female students at the university that have a job outside the university.
    (d) The proportion in the sample ($\frac{19}{67}$).
    (e) No, it's a survey (a type of observational study).
    (f) No, it's a voluntary response sample.
    (g) Possibly, but there could be all sorts of bias introduced. Students who work may very well be more or less likely to send in their survey than those who do not.

16. (a) All black bear gall bladders in Ontario.
    (b) 22 grams is a characteristic of a sample, and is thus a statistic.
    (c) 24 grams is a characteristic of the entire population, and is thus a parameter.
    (d) Something else. It's more of a sample of convenience (the next 3 bears encountered), which could possibly be extremely biased.

17. No, they do not represent a simple random sample. Values near the middle are more likely than those on the ends, and thus middle sized sunflowers are more likely to be selected than small or large sunflowers.

18. (a) False. Clearly the survey gives *some* information. As it relates to the opinions of members of this organization, it would yield valuable information.
    (b) False. This type of delay would not be a big problem in this situation.
    (c) False. A sample of 912 is a fairly large sample, and can provide valuable information.
    (d) True. The population of interest is Americans in general, but only the opinions of the members of this pro-gun organization are reported.

19. (a) All third graders in the school.
    (b) It is a characteristic of a sample, and is therefore a statistic.
    (c) No. The extent of the bias that might be present in this type of sample is difficult to assess, but it's not a simple random sample. (Not all possible samples are equally likely under this sampling scheme.)

20. (a) False. The sampling design was not stated, but it is highly unlikely that the sample represents a simple random sample.
    (b) False. The wording is pretty reasonable, and would not cause much bias.
    (c) False. The value 72% came from a sample, and is thus a statistic.
    (d) False. A sample of size 1500 is quite a large sample for most practical purposes.

21. (a) It is an experiment (the researchers imposed a condition (the aspirin) on some of the patients).
    (b) Whether or not the patient had a heart attack.
    (c) Whether or not the patient received aspirin.
    (d) No. It is sometimes desirable, but not necessary.
    (e) False. This is not what confounding means.

22. (a) No, the conclusions would not be as strong. Here we are discussing an observational study, and thus we could not reach conclusions of a causal nature. The experiment in Question 21 could give strong evidence of a causal link.
    (b) Physicians might not want to respond if they've had a heart attack. Or perhaps they would be more likely to respond if they've had a heart attack, as they may wish to contribute to research of this nature. (We can suspect possible biases without knowing the direction or magnitude.) And, of course, any physician who previously died of a heart attack would be unable to respond.

23. (a) False, there are 2 treatments groups (shock/no shock). (What we might call a *treatment* group and a *control* group.)
    (b) False, it is an experiment.
    (c) False. Experiments can give strong evidence of a causal effect.
    (d) False. The response variable is the number of matching pairs.

24. (a) Task completion time.
    (b) The harness variable (whether or not a worker is wearing a harness).
    (c) It is an experiment (the researchers imposed an intervention, the harness, on some workers).
    (d) Yes. It appears to be a reasonable experimental design, and as such can provide evidence of a causal link.
    (e) Without the randomization, bias can be introduced. Perhaps workers who felt they might be slowed down a great deal by the harness would choose to not wear one. Or perhaps faster workers would choose to not wear a harness. There is a variety of possible biases.

25. (a) **3**3**69**6 9**154**4 76248. The first 3 numbers selected are 3, 6, and 9. So Jerry, Bo, and Eddie go to the control group. The second set of 3 numbers is 1, 5, 4, so Pete, Renaldo, and Alphonse go to the cheese group. The others (2, 7, 8 – Tom, Little Pete, and Huey) go to the electric shock group.
    (b) This may introduce bias. You might want to save Little Pete because he is too weak and frail, or a variety of other reasons that might introduce bias.

26. There appears to be a relationship, but we cannot say it is due to the yoga. Perhaps women who practice yoga take care of themselves and are simply healthier overall. Based on this type of survey, we have little evidence that yoga causes improved

health.

27. a: An unmeasured variable that is related to both the response and explanatory variables.

28. (a) Reduction in blood pressure.
    (b) The types of drug.
    (c) It is a randomized experiment.
    (d) Yes. A well designed experiment can give strong evidence of a causal link.

29. Only (a) is true. This observational study does not give strong evidence of a causal link. There may be a causal link, but this study certainly doesn't show it. Note that those who are more likely to get into an accident may be more likely to purchase this service.

30. (a) The bonding agents.
    (b) The bonding agents.
    (c) It is an experiment.
    (d) Yes. This type of experiment can show strong evidence of a causal link.

31. (a) An observational study (a study of coroner records, for example).
    (b) If the data is available, an observational study would be best. (Perhaps records from various companies are available.) This may not be possible, and a survey of workers may have to suffice.
    (c) An experiment.
    (d) An experiment would not be possible here. And it would be tough to get appropriate data for an observational study. We would need to rely on a survey. But unfortunately, even the survey results may not be very reliable, as people may be likely to lie when faced with questions about cocaine use or marital infidelity.
    (e) An observational study (a study of coroner records, for example).

32. (a) The experimental surgery. The results from the experiment are more reliable (less subject to bias).
    (b) The medical condition of the dog. Owners of dogs that are in very poor shape may be more willing to gamble on the experimental surgery.

Chapter 3

Descriptive Statistics

J.B.'s strongly suggested exercises: 1, 4, 7, 8, 9, 10, 11, 16, 17, 20, 21, 22, 24, 25, 36, 46, 49

3.1 Introduction

3.2 Plots for Qualitative and Quantitative Variables

3.2.1 Plots for Qualitative Variables

1. (a) Most victims (67%) were acquaintances of the male perpetrator. Slightly less than one quarter (24%) were the male murderer's intimate partner. Small percentages of the victims were family members (4%) or strangers (4%).
   (b) Compared to the victims of female murderers, victims of male murderers are more likely to be acquaintances and less likely to be intimate partners or family members.
   (c) No, of course not. This is sample data, and for sample data there is *always* the chance that what we observe is simply due to chance and not a real effect. Later on, we will conduct a χ^2 test on this data, in order to assess how much evidence there is of a real effect.
   (d) The proportions in each category are:
   - Acquaintances: $\frac{61}{91} \approx .670$
   - Partners: $\frac{22}{91} \approx .242$
   - Family: $\frac{4}{91} \approx .044$
   - Stranger: $\frac{4}{91} \approx .044$

These proportions make up the areas of the sections of the pie chart. The pie chart for this data is illustrated in Figure 3.1.

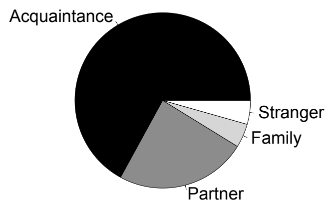


Figure 3.1: Relationship to the victim for male murderers in Finland.

3.2.2 Plots for Quantitative Variables

3.3 Numerical Measures

3.3.1 Summation Notation

2. (a) 22 (the third number in the list).
   (b) $\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3 = 8 + 14 + 22 = 44$.
   (c) $\sum x_i$ is shorthand notation for $\sum_{i=1}^{n} x_i$. $\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + x_4 = 8 + 14 + 22 + (-5) = 39$.
   (d) $\sum x_i^2 = 8^2 + 14^2 + 22^2 + (-5)^2 = 769$.

3.3.2 Measures of Central Tendency

3.3.2.1 Mean, Median, and Mode

3. (a) $\bar{x} = \frac{\sum x_i}{n} = \frac{1 + 5 + 2 + (-3) + 987}{5} = 198.4$.
   (b) When ordered from smallest to largest, the 5 observations are: -3, 1, 2, 5, 987. The median is 2, the middle value.
   (c) No observation occurs more often than any other (they all occur once), and so there is no mode.

(d) The extreme value is 987. When this value is removed, $\frac{\sum x_i}{n} = \dfrac{1+5+2+(-3)}{4} = 1.25$.

(e) The ordered values are: -3, 1, 2, 5. The median is the average of the two middle values: $\frac{1+2}{2} = 1.5$.

4. (a) $\bar{x} = \frac{2870+2620+3120+3620}{4} = 3057.5$.

(b) The ordered values are: 2620, 2870, 3120, 3620. The median is the average of the two middle values: $\dfrac{2870+3120}{2} = 2995$.

(c) Grams (the same units as the variable).

(d) Grams (the same units as the variable).

(e) Either one is perfectly acceptable, and it's fine to report both. (There are no extreme values that might make the mean a misleading measure of centre.)

3.3.2.2 Other Measures of Central Tendency

5. The trimmed mean is not as influenced by extreme values as the untrimmed mean. This can be advantageous in certain situations, as we don't want a single value to have undue influence on our statistics and conclusions. A disadvantage is that the trimmed mean omits some possibly valuable information.

3.3.3 Measures of Variability

6. (a) $6.5, -3.5, -8.5, 5.5$. The deviation for the ith observation is $x_i - \bar{x}$, where $\bar{x} = 11.5$. $x_1 - \bar{x} = 18 - 11.5 = 6.5$, $x_2 - \bar{x} = 8 - 11.5 = -3.5$, $x_3 - \bar{x} = 3 - 11.5 = -8.5$, $x_4 - \bar{x} = 17 - 11.5 = 5.5$.

(b) 0. The deviations *always* sum to 0.

7. (a) Range = Maximum − Minimum = 3620 − 2620 = 1000.

(b) MAD $= \dfrac{\sum |x_i - \bar{x}|}{n}$, where $\bar{x} = 3057.5$.

MAD $= \dfrac{|2870 - 3057.5| + |2620 - 3057.5| + |3120 - 3057.5| + |3620 - 3057.5|}{4}$

MAD $= \frac{187.5+437.5+62.5+562.5}{4} = 312.5$ grams.

(c) $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$, where $\bar{x} = 3057.5$.

$s^2 = \dfrac{(2870-3057.5)^2 + (2620-3057.5)^2 + (3120-3057.5)^2 + (3620-3057.5)^2}{4-1} = 182291.7$.

(d) $s = \sqrt{182291.7} = 426.9563$.

(e) grams^2 (the square of the units of the variable).
(f) grams (the units of the variable).

8. (a) The 4 values must be: $0, 0, 500, 500$.
   (b) These 4 values have a standard deviation of $s = 288.6751$.

9. (a) Any 4 values that are equal will work (e.g. 0,0,0,0, or 75, 75, 75, 75).
   (b) If the 4 values are equal, $s = 0$.

10. (a) True. (The standard deviation will be greater than the variance if $0 < s < 1$.)
    (b) False. ($s \geq 0$.)
    (c) True. (The mean is a measure of location, but the standard deviation is a measure of variability. There is no general relationship between them.)
    (d) True. (Q_3 is a measure of position, but the standard deviation is a measure of variability. There is no general relationship between them.)
    (e) False. (The standard deviation cannot be less than the average distance from the mean.)

3.3.3.1 Interpreting the standard deviation

11. (a) Yes (the empirical rule would apply), since the distribution is roughly mound-shaped.
    (b) Approximately 68% of the observations would lie within 7.2 of 22.4 (within 1 standard deviation of the mean).
    (c) Approximately 95% of the observations would lie within 14.4 of 22.4 (within 2 standard deviations of the mean).
    (d) All or almost all of the observations would lie within 21.6 of 22.4 (within 3 standard deviations of the mean). (At most a very small proportion of observations would fall outside of that range.) For this particular data set, no observations fall outside of that range of values.

12. (a) Yes, Chebyshev's theorem would apply. Chebyshev's theorem applies to *every* data set.
    (b) Nothing. Chebyshev's theorem is only informative if $k > 1$.
    (c) At least 75% of the observations would lie within 14.4 of 22.4 (within 2 standard deviations of the mean).
    (d) At least 88.9% of the observations would lie within 21.6 of 22.4 (within 3 standard deviations of the mean).

13. (a) The empirical rule would not officially apply to this data set, since we would not call this a mound-shaped distribution (it's skewed to the right). However, the empirical rule still works pretty well here. For the data given in the plot,

69.6% of the observations fall within 1 standard deviation of the mean, 95.2% of the observations fall within 2 standard deviations of the mean, and 99.1% of the observations fall within 3 standard deviations of the mean. So the empirical rule can still be informative, even if there is some skewness.

(b) Yes, of course Chebyshev's theorem would apply, since Chebyshev's theorem applies to *every* data set.

3.3.3.2 Why divide by $n-1$ in the sample variance formula?

14. Dividing by $n-1$ results in an estimator that, on average, equals the population variance σ^2. If we divided by n, the resulting estimator would tend to be too small.

15. 86. Since we do not know the population mean μ, we will need to estimate it with $\bar{x}$, resulting in a loss of 1 degree of freedom. In this type of scenario, the degrees of freedom are $n-1$.

3.3.4 Measures of Relative Standing

3.3.4.1 Z-scores

16. (a) $z_1 = \frac{x_1-\bar{x}}{s} = \frac{2870-3057.5}{426.9563} = -0.439$. $z_2 = \frac{x_2-\bar{x}}{s} = \frac{2620-3057.5}{426.9563} = -1.025$. $z_3 = \frac{x_3-\bar{x}}{s} = \frac{3120-3057.5}{426.9563} = 0.146$. $z_4 = \frac{x_4-\bar{x}}{s} = \frac{3620-3057.5}{426.9563} = 1.317$.
    (b) 0. The mean of all z scores is always 0.
    (c) 1. The standard deviation of all z scores is always 1.
    (d) The baby's weight is very large relative to other babies in the data set. (The baby's weight is 4.6 standard deviations greater than the mean.)
    (e) The baby's weight is a just a little small relative to other babies in the data set. (The baby's weight is 0.4 standard deviations less than the mean.)

17. (a) True.
    (b) True.
    (c) False. Todd's score was better than almost all writers of the test.
    (d) True.

3.3.4.2 Percentiles

18. (a) 691. The ordered weights of the 8 boxes of cereal are:

$$684, 684, 684, 686, 686, 691, 691, 691$$

We need to find the 80th percentile. $np = 8 \times 0.80 = 6.4$. We round up to the next largest integer (7), and the 80th percentile is the 7th ranked value (691).

(b) 684. We need to find the 25th percentile. $np = 8 \times 0.25 = 2$. The 25th percentile is the average of the 2nd and 3rd ranked values: $\frac{684+684}{2} = 684$.

19. 171 cm. (171 is the only plausible value. 143 cm and less are far too small, and 200 cm is far too big.) **?** estimate the 90th percentile of women 20 years and over in the United States to be 170.9 cm.

3.4 Boxplots

20. (a) Approximately 35.
    (b) Approximately 80.
    (c) 4 (Sample A and Sample B each have 2).
    (d) A little greater than 20.

21. (a) Approximately 74 mm.
    (b) Approximately 66 mm.
    (c) The variances of the two distributions appear to be very similar. The distribution for males appears to be shifted higher than that of females. (The males of this species seem to have a greater tail length on average.)

3.5 Linear Transformations

22. (a) $\bar{x} = 20 \times (100 + 5) = 2100$.
    (b) $s^2 = 20^2 \times 20^2 = 160000$. (Note that the additive constant (5) does not change the variance.)
    (c) $s = 20 \times 20 = 400$.
    (d) $20 \times (90 + 5) = 1900$.

23. All of the measures of *location* or *position* (mean, median, Q_3) would change. None of the measures of *variability* (IQR, variance, standard deviation) would change.

3.6 Chapter Exercises

3.6.1 Basic Calculations

24. (a) $\bar{x} = 41.25$.
    (b) Median = 30.
    (c) $s^2 = 2471.583$.
    (d) $s = 49.71502$.
    (e) 109.

25. (a) $\bar{x} = 8.214286$.
    (b) Median = 8.4. The ordered values are $-1.7, 3.1, 8.2, 8.4, 8.8, 9.6, 21.1$. The median is the fourth number in this ordered list.
    (c) $Q_3 = 9.6$. By the method given in the text: $7 \times 0.75 = 5.25$, so we round up to 6 and choose the 6th value in the ordered list.
    (d) $IQR = Q_3 - Q_1 = 9.6 - 3.1 = 6.5$. By the method given in the text: $7 \times 0.25 = 1.75$, so we round up to 2 and Q_1 is the 2nd value in the ordered list. We found Q_3 in 25c.
    (e) $s^2 = 48.7981$.
    (f) $s = 6.985563$
    (g) Yes, 21.1 is an outlier. (21.1 falls outside of the range $Q_1 - 1.5 \times IQR = -6.65$ to $Q_3 + 1.5 \times IQR = 19.35$.)

26. (a) $\bar{x} = 30$.
    (b) Median = 25. The ordered list is: $1, 4, 18, 32, 36, 89$. Since there are 6 observations, the median is the average of the 3rd and 4th values in the ordered list. (Median = $\frac{18+32}{2}$.)
    (c) $Q_1 = 4$. By the method in the text: $6 \times 0.25 = 1.5$, so we round up to 2 and Q_1 is the second value in the ordered list.
    (d) $IQR = Q_3 - Q_1 = 36 - 4 = 32$. ($Q_1 = 4, Q_3 = 36$.) We found Q_1 in 26c. We can find Q_3 using the same approach: $6 \times 0.75 = 4.5$, so we round up to 5 and Q_3 is the fifth value in the ordered list.
    (e) $s^2 = 1036.4$.
    (f) $s = 32.19317$.

27. (a) $\bar{x} = 149.2$.
    (b) Median = 6.
    (c) $s = 331.5316$.
    (d) $s^2 = 109913.2$.
    (e) If 742 is thrown out: $\bar{x} = 1$. Note that the mean has decreased a great deal.
    (f) If 742 is thrown out: Median = 1. Note that the median has changed by a

much smaller amount than the mean.

(g) If 742 is thrown out: $s = 11.37248$. Note that the standard deviation decreased a great deal when the outlier was discarded.

28. (a) The values are $-1.7, -1.1, -1.1, -1.0, -.7, -.6, -.6, -.3, -.2, 0.0, 0.3, 0.3, 0.3, 0.5, 1.0, 1.1, 1.3, 1.8, 1.9, 2$
They result in a mean of $\bar{x} = 0.18$.
(b) Median $= 0.15$.
(c) $s^2 = 1.264842$.
(d) $s = 1.124652$.
(e) Range $= 2.4 - (-1.7) = 4.1$.

29. (a) Yes, the mean is greater than the median. (The large values in the right tail affect the mean more than the median.)
(b) Yes, the distribution shows some right skewness (it is stretched out to the right).
(c) The fifth value in the five-number summary is the maximum. The first value in the five number summary is the minimum. Max − Min $= 5.8 - 0.0 = 5.8$.

30. (a) The values are $-4.9, -4.7, -4.5, -3.9, -3.9, -3.4, -3.1, -3, -1.6, 1.1, 2.1, 2.3, 3.2, 3.4, 3.9$.
These result in a mean of $\bar{x} = -1.133333$.
(b) Median $= -3$.
(c) Since we would in effect be multiplying every value by $\frac{1}{10}$, every one of the listed statistics would change. All but the variance would change by a factor of $\frac{1}{10}$. The variance would change by a factor of $(\frac{1}{10})^2 = \frac{1}{100}$.

3.6.2 Concepts

31. (a) Approximately 12.
(b) Approximately 2.
(c) Approximately 25.
(d) No observations are very extreme.
(e) True. The distribution is approximately symmetric, so the values of the mean and median would be close.
(f) Boxplot A.

32. (a) Since the deviations must sum to 0 ($\sum(x_i - \bar{x}) = 0$), the fifth deviation must be -1.9.
(b) This is impossible to determine; the mean could have been any value.
(c) $s = 2.475884$. $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{(-2.3)^2+3.7^2+1.2^2+(-0.7)^2+(-1.9)^2}{5-1} = 6.13$. $s = \sqrt{6.13} = 2.475884$.

33. The data set might be $12, 12, 12, 12, 22$ or $134, 134, 134, 134, 144$, or anything else of this form. The variance of any data set of this form is the same: $s^2 = 20$.

34. (a) True, the median is definitely less than 4. (It would be close to 1.)
    (b) False. The distribution is skewed to the right, and the mean would be greater than the median.
    (c) False. (Very few observations in this data set are greater than 5, but 25% of the values are greater than Q_3.)
    (d) True.
    (e) False. The distribution is right-skewed.
    (f) False. For a right-skewed distribution such as this one, Q_3 would lie farther above the median than Q_1 lies below it.
    (g) This is debatable. There are some large values that could be considered outliers.

35. (a) Boxplot A, as it shows the right-skewness of the histogram.
    (b) Boxplot C.
    (c) Boxplot B, which has an interquartile range of approximately 1.2.
    (d) Boxplot A and Boxplot C contain the same number of outliers.

36. (a) The mean, variance, and standard deviation would change (they would increase). The median would not change, nor would Q_1, Q_3 or the IQR.
    (b) The mean, median, and IQR would not change. The standard deviation and variance would increase.

37. (a) True. If the standard deviation is 0, all the values in the data set must be equal. This implies all the values in the 5 number summary are equal.
    (b) True, since all the deviations are 0.
    (c) True. (The standard deviation and variance are both equal to 0.)
    (d) True. (The mean and median are both equal to 0.)
    (e) True. In the new data set, there would be 99 values that are equal and 1 that is bigger. The IQR, minimum, and median would still be 0.

38. $s = 0$. If the standard deviation of the 4 number data set is 0, that means all values must equal 50. Since the added value is also 50, then all numbers in the data set are equal and $s = 0$.

39. (a) False. The standard deviation can be greater or less than the median—they are measuring completely different things.
    (b) False. The variance is usually greater than the standard deviation, but not if $0 \leq s^2 \leq 1$.
    (c) False. The mean can be greater or less than the standard deviation—they are measuring completely different things.

(d) True. The standard deviation is a little greater than the *average distance from the mean*, which has a maximum of 1 in this case. The standard deviation of the sample (0, 2) for example, is $s = 1.414214$.

(e) True. The median is *usually* less than Q_3, but they can be equal if there are repeated values in the data set. For example, if the data set is $0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 12, 24$, then the median and Q_3 will both equal 0.

40. (a) True. It would not be possible in this case for the distribution to the "stretched out" to the left.
    (b) True. The units of the variance are the square of the units of the variable.
    (c) False. (By the same logic that tells us that Q_1 is not in general exactly half of the median.)
    (d) False. Any of the values in the five-number summary can be negative. (Recall that the five-number summary is: Minimum, Q_1, Median, Q_3, Maximum.)
    (e) False. Outliers increase both the standard deviation and variance (and sometimes increase them a great deal).

41. (a) False. If the distribution is right-skewed, the mean will be greater than the median.
    (b) False. These two statistics are measuring completely different things. Symmetry would not make them equal.
    (c) False. All or almost all observations will fall within 5 standard deviations of the mean.
    (d) False.
    (e) False. Although values less than Q_1 would usually have z-scores that are negative, it is not impossible for a value less than Q_1 to be greater than the mean. (e.g. consider the data set: $-328471934871389 47, 1, 1, 2, 3, 3, 8, 48, 52, 113$.)
    (f) True. If the range is 0, then all values in the data set must be equal. This implies the variance and IQR will also equal 0.

42. (a) The median will equal 15, regardless of what those 2 values are.
    (b) It is impossible to determine the value of the mean.
    (c) It is impossible to determine the value of the variance.

43. (a) If we let x represent the original variable, and x^* represent the transformed variable, then we need to find a and b where $x^* = a + bx$. The standard deviation changed from 5 to 1, and thus the multiplicative constant must be either $b = \frac{1}{5}$ or $b = -\frac{1}{5}$. We can now solve for a in either case. The two possible transformations that satisfy the conditions are $x^* = 12 + \frac{1}{5}x$ and $x^* = 18 - \frac{1}{5}x$.
    (b) Carrying out the transformation $x^* = 12 + \frac{1}{5}x$, $x^* = 12 + \frac{1}{5}10 = 14$.

44. The possible values are 52 and 352. Three consecutive digits have a sample standard deviation of 1 (e.g. $s = 1$ for the values $5, 6, 7$). Three values that are increasing by

100 (e.g. 500, 600, 700) will have a standard deviation of $s = 100$. So the missing value could be either 52 or 352. (Verify this on your calculator or computer.)

3.6.3 Applications

45. (a) Just a little under 70.
    (b) Somewhere in the high 50s.
    (c) Approximately 20.
    (d) It is tough to say with any certainty, but there is a touch of left-skewness, implying the mean would likely be less than the median.
    (e) The range is approximately 80. The standard deviation will likely fall somewhere in or close to the interval $\frac{80}{6}$ to $\frac{80}{4}$, or somewhere in the neighbourhood of 13 to 20. (For mound-shaped distributions, the standard deviation often falls within or close to the interval $\frac{\text{Range}}{6}$ to $\frac{\text{Range}}{4}$.)

46. (a) The distribution is left skewed, with a clump of outliers at 0.
    (b) It would be more meaningful to report the average grade *for students that handed in the assignment*, as well as the percentage of students that handed it in. For example, we might report that 94% of students handed in this assignment, and those students had an average grade of 20.0 out of 25.

47. (a) $\bar{x} = \frac{\text{Total number of eggs}}{\text{Total number of scars}} = \frac{142 \cdot 1 + 194 \cdot 2 + 9 \cdot 314 \cdot 4}{142 + 194 + 9 + 14} = 1.7075.$
    (b) 2. There are 359 observations, so the median is the 180th ranked value. There are 142 ones and 194 twos in the data set, so the median must be 2.
    (c) $s^2 = \frac{142(1-1.7075)^2 + 194(2-1.7075)^2 + 9(3-1.7075)^2 + 14(4-1.7075)^2}{359-1} =$ 0.49243.
    $s = \sqrt{0.49243} = 0.7017.$

48. (a) The distribution is slightly skewed to the left.
    (b) The 90th percentile would be a little less than 60 kgf.
    (c) There is no legitimate way to estimate this quantity from the given data. It we were to start playing guessing games, we could guess that the 90th percentile for adult Germans would likely be less than the 90th percentile of this data set (since the sample was made up of males of university age).
    (d) Both the mean and median would be reasonable measures of central tendency here. It would be perfectly fine to report either or both.
    (e) From the boxplot, the median looks to be approximately 52.5. Since the distribution has slight left skewness, the mean is likely slightly less than this quantity. (Based on the raw data used to create these plots: Median = 52.25 and $\bar{x} = 50.91$.)

(f) While the distribution shows slight left skewness, it is roughly mound-shaped, so the empirical rule provides a bit of guidance. The standard deviation likely falls in or close to the interval $\frac{\text{Range}}{6}$ to $\frac{\text{Range}}{4}$. The range is somewhere between 30 and 35 kgf, so let's estimate the range to be 33. The standard deviation is likely in or near the interval $\frac{33}{6} = 5.5$ to $\frac{33}{4} = 8.25$. (Based on the raw data used to create these plots: $s = 7.74$.)

49. (a) The distribution looks roughly symmetric, with one very large outlier at approximately 120%.
    (b) The outlier looks to be approximately 85 units above the mean of the remaining data. Removing the outlier would result in a decrease in the mean of approximately $\frac{85}{23}$. The estimated mean without the outlier is approximately $36.6 - \frac{85}{23} = 32.90$. (If we return to the raw data and remove the outlier, $\bar{x} = 32.89$.)
    (c) We should be extremely reluctant to remove a real observation. When possible, we should check to make sure that an extreme observation was correctly recorded and was not in error. If the value is correct, then we should never simply ignore it. We should be more inclined to use measures that are not as influenced by extreme values (such as reporting the median instead of the mean). Reporting the results both with and without the outlier is another option.
    (d) It depends on the reason why the machine did not record the break. If it was simply a random occurrence that had nothing to do with the elongation of the cord, then this is not a problem. If the reason for the malfunction was related to the elongation, then this might be a big problem. For example, if the machine tends to fail to record the break when the elongation percentage is very large, then ignoring these data points might incorporate very strong bias (our reported mean and median will be too small).

50. (a) From the boxplot, the median looks to be approximately 27 grams. The plots do not show any major outliers or skewness, so the mean and median are likely close in value (27 grams would also be a reasonable estimate of the mean). (The mean and median of the raw data are both 27.2 grams.)
    (b) It depends entirely on how this sample was drawn, but the sample is very possibly strongly biased. In practice, it is impossible to get a simple random sample of crayfish from a lake. It is very likely that the method of capturing the crayfish introduced bias.

51. (a) Since this is just a linear transformation, we can carry out the transformation on the mean. The new mean is $\bar{x}^* = \frac{47.5975-40}{2.54} = 2.991142$.
    (b) Similarly for the median: $\frac{47.12-40}{2.54} = 2.803150$.
    (c) Subtracting 40 will not change the measures of variability, and thus only the division by 2.54 matters here. $s_{x^*} = \frac{1}{2.54} \times 1.212878 = 0.477511$.

(d) $s_{x^*}^2 = 0.477511^2 = 0.2280168$.

52. A linear transformation is required: $\bar{x}^* = 2.20(x + 2) = 4.4 + 2.2x$. The mean in pounds is $\bar{x}^* = 4.4 + 2.2 \times 15.10 = 37.62$. The additive constant does not affect the standard deviation, and the new standard deviation is $s_{x^*} = 2.2 \times 1.70 = 3.74$.

Chapter 4

Probability

J.B.'s strongly suggested exercises: 2, 4, 5, 6, 7, 8, 14, 16, 17, 21, 30, 35, 39, 43, 49, 57, 61

4.1 Introduction

4.2 Basics of Probability

4.2.1 Sample Spaces and Sample Points

1. (a) Depending on our needs, we might define the sample space in different ways. One possibility is the set of all $52 \times 51 = 2652$ orderings of two cards: $S = \{(2c, 3c), (2c, 4c), \ldots, (As, Ks)\}$.
   Another possibility is the set of all pairs of cards, without regard to order. (There are 1326 possible pairs of cards, if order does not matter.)
   (b) Under the first definition of the sample space given in 1a there are $52 \times 51 = 2652$ sample points: $(2c, 3c)$, $(2c, 4c)$, $\ldots$, (As, Ks).
   (c) Yes, any two card ordering has the same chance occurring.
   (d) $4 \times 3 = 12$ of the sample points are a pair of twos. $P(\text{Pair of twos}) = \frac{12}{2652} = \frac{1}{221}$.
2. (a) The sample space is the set of the $6 \times 6 = 36$ possible outcomes: $S = \{(1,1), (1,2), \ldots, (6,6)\}$.
   (b) There are $6 \times 6 = 36$ sample points: $(1,1), (1,2), \ldots, (6,6)$. (Where $(1,2)$ represents rolling a 1 on the first die and a 2 on the second.)

(c) Yes, each of the sample points has a $\frac{1}{36}$ chance of occurring.
(d) Only one simple event (1,1) results in a sum of 2. $P(\text{Sum is 2}) = \frac{1}{36}$.
(e) Six sample points result in a sum of 7: $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$. $P(\text{Sum is 7}) = \frac{6}{36} = \frac{1}{6}$.

4.3 Rules of Probability

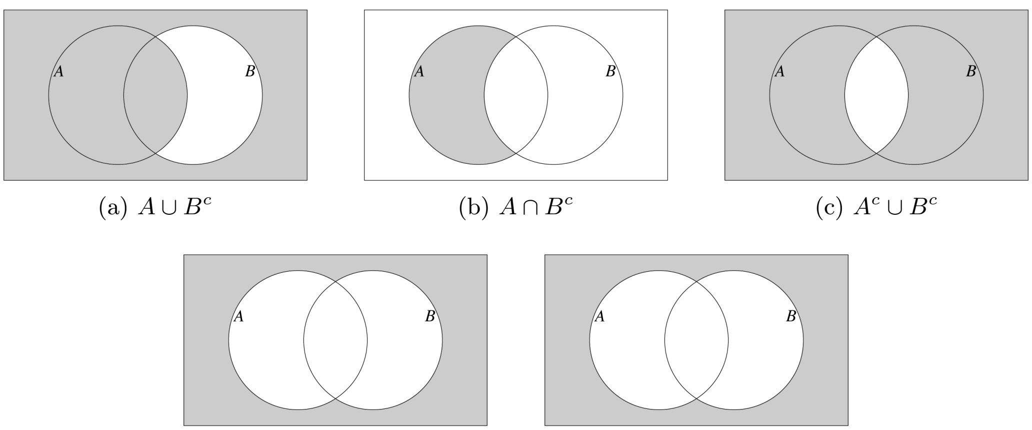
The Intersection of Events

Mutually Exclusive Events

The Union of Events

Complementary Events

3. See Figure 4.1.



(a) $A \cup B^c$ (b) $A \cap B^c$ (c) $A^c \cup B^c$

(d) $A^c \cap B^c$ (e) $(A \cup B)^c$

Figure 4.1: Venn diagrams for Question 3.

4. This scenario is illustrated in Figure 4.2.

(a) $P(A^c) = 1 - P(A) = 1 - 0.6 = 0.4$.
(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.3 - 0.2 = 0.7$.
(c) $P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.7 = 0.3$.
(d) No. $P(A \cap B) \neq 0$, so A and B are not mutually exclusive.

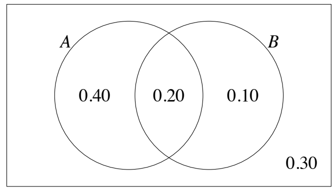


Figure 4.2: Venn diagram for Question 4.

(e) This one is a little harder to visualize. $A^c \cap B^c = (A \cup B)^c$ (drawing a Venn diagram helps to visualize this—see Figure 4.11). So $P(A^c \cap B^c) = 0.3$, and A^c and B^c are therefore not mutually exclusive.

5. (a) $D \cap E^c$. $P(D \cap E^c) = P(\text{O negative}) = 0.07$ (this can be read straight from the table).
   (b) $P(\text{O}) = 0.39 + 0.07 = 0.46$. $P(\text{Rh positive}) = 0.36 + 0.076 + 0.025 + 0.39 = 0.851$.

$$\begin{aligned} P(\text{O} \cup \text{Rh positive}) &= P(\text{O}) + P(\text{Rh positive}) - P(\text{O} \cap \text{Rh positive}) \\ &= 0.46 + 0.851 - 0.39 = 0.921. \end{aligned}$$

   (c) The ABO blood groups are mutually exclusive (a person is classified into one and only one of A, B, AB, or O), so the following events are mutually exclusive: $(A, B), (A, C), (A, D), (B, C), (B, D), (C, D)$.
   (d) $(A \cap E^c) \cup (D \cap E^c)$, or $(A \cup D) \cap E^c$. $P((A \cap E^c) \cup (D \cap E^c)) = 0.06 + 0.07 = 0.13$.

Conditional Probability

Independent Events

6. (a) The addition rule tells us that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, so $0.65 = 0.60 + 0.20 - P(A \cap B)$, which implies $P(A \cap B) = 0.15$.
   (b) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.15}{0.20} = 0.75$.
   (c) $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.15}{0.60} = 0.25$.
   (d) No, since $P(A|B) \neq P(A)$ $(0.75 \neq 0.60)$.

7. (a) $P(B|C) = 1$. If we are *given* that the number on the top face is a 6 (event C), we know with certainty it is an even number (event B). Alternatively, we could use the conditional probability formula:

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{1/6}{1/6} = 1$$

($P(B \cap C) = \frac{1}{6}$ because the intersection of B and C is 6, one of the six possible equally likely outcomes.)

(b) $P(C|B) = \frac{1}{3}$. If we are given the number on the top face is an even number (event B – the numbers 2, 4, 6), then there is a $\frac{1}{3}$ probability it is a 6 (event C). Alternatively, we could use the conditional probability formula:

$$P(C|B) = \frac{P(B \cap C)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

($P(B \cap C) = \frac{1}{6}$ because the intersection of B and C is the number 6.)

(c) $P(C|B^c) = 0$. If we are given the number is *odd* (event B^c, since B is the even numbers), then the number cannot possibly be a 6. Alternatively, we could use the conditional probability formula:

$$P(C|B^c) = \frac{P(B^c \cap C)}{P(B^c)} = \frac{0}{3/6} = 0$$

($P(B^c \cap C) = 0$ because B^c and C have no events in common.)

(d) $P((B \cap C)|A) = \frac{1}{3}$. $B \cap C = \{6\}$ (the only event B and C share is 6). If we are given A has occurred (one of 4, 5, 6 has occurred), then the probability a 6 has occurred is $\frac{1}{3}$. Alternatively, we could use the conditional probability formula:

$$P((B \cap C)|A) = \frac{P(B \cap C \cap A)}{P(A)} = \frac{1/6}{3/6} = \frac{1}{3}$$

($P(B \cap C \cap A) = \frac{1}{6}$ because the intersection of B and C and A is the number 6.)

(e) None of the pairs are independent. $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{6}$.

$$P(A|B) = \frac{2}{3} \neq P(A) \implies \text{A and B are not independent}$$
$$P(A|C) = 1 \neq P(A) \implies \text{A and C are not independent}$$
$$P(B|C) = 1 \neq P(B) \implies \text{B and C are not independent}$$

The Multiplication Rule

8. (a)

$$\begin{aligned} P(\text{Both cards are hearts}) &= P(\text{First card is a heart}) \times P(\text{Second card is a heart} \mid \text{First card is a heart}) \\ &= \frac{13}{52} \times \frac{12}{51} \\ &= \frac{156}{2652} = \frac{1}{17} \end{aligned}$$

(b)

$$\begin{aligned}
P(\text{Both cards are fives}) &= P(\text{First card is a 5}) \times P(\text{Second card is a 5} \mid \text{First card is a 5}) \\
&= \frac{4}{52} \times \frac{3}{51} \\
&= \frac{12}{2652} = \frac{1}{221}
\end{aligned}$$

(c) $\frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} = \frac{17160}{6497400} \approx 0.00264.$

9. (a)

$$\begin{aligned}
P(\text{Both dice are fours}) &= P(\text{First die is a four}) \times P(\text{Second die is a four} \mid \text{First die is a four}) \\
&= \frac{1}{6} \times \frac{1}{6} \text{ (The rolls are independent.)} \\
&= \frac{1}{36}
\end{aligned}$$

(b) Since the rolls are independent, these two events are independent, and we can simply multiply the two probabilities together:

$$P(\text{Roll an even number, then a number greater than 4}) = \frac{3}{6} \times \frac{2}{6} = \frac{6}{36} = \frac{1}{6}$$

4.4 Examples

10. We need to find $P(A|B) = \frac{P(B \cap A)}{P(A)}$. $P(A)$ is given in the question, so we just need to find $P(B \cap A)$. We *cannot* say that $P(B \cap A) = P(B)P(A)$, since we do not know if A and B are independent (in fact, since $P(B|A) \neq P(B)$, we know that A and B are *not* independent). We need to find the probability of the intersection in another way. What is the other bit of information we are given in the question? We are given that $P(B|A) = \frac{P(B \cap A)}{P(A)} = 0.25$, which implies $P(B \cap A) = 0.25 \times 0.80 = 0.20$. Now we can calculate:

$$P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{0.20}{0.40} = 0.5$$

11. (a) See Figure 4.3.
(b) $P(A \cup B) = 0.42 + 0.44 - 0.16 = 0.70$.
(c) $P(A \cup B \cup C) = 0.82$.
(d) $P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{0.04}{0.12} = \frac{1}{3}$.

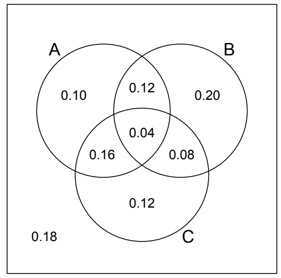


Figure 4.3: Venn diagram for Question 11a.

12. (a) Since any one of these 1000 individuals is equally likely to be picked, this probability is simply the number of Republicans over the total number of individuals:
$$P(\text{Republican}) = \frac{60 + 270}{1000} = 0.330$$
(b) This is the number of individuals that said they approve, over the total number of individuals:
$$P(\text{Approve}) = \frac{60 + 361 + 139}{1000} = 0.560$$
(c) The conditional probability formula can be used here, but we can also answer this question quickly using a little bit of logic. We are given that the individual is a Republican, which means we are dealing with only the 330 Republicans in the sample. *Of the 330 Republicans*, 60 approve, and so $P(\text{Approve}|\text{Republican}) = \frac{60}{330}$. We could have used the conditional probability formula had we so desired:
$$P(\text{Approve}|\text{Republican}) = \frac{P(\text{Approve} \cap \text{Republican})}{P(\text{Republican})} = \frac{60/1000}{330/1000} = \frac{60}{330}$$
Note that P(Approve) $\neq$ P(Approve|Republican). This means the events: *The person is a Republican*, and *The person said they approve of the way Obama is handling his job as president* are not independent. (Which should make sense, if you know a little about United States politics.)

13. (a) A and B are independent in Plot B. (We can't determine independence with certainty by just looking at the plot, without the information that A and B are independent for one of the plots. It can only be plot B. This plot was created such that $P(A) = 0.1$, $P(B) = 0.2$, and $P(A \cap B) = 0.02$.)
(b) Plot A: $P(A|B) < P(A)$.
Plot B: $P(A|B) = P(A)$.
Plot C: $P(A|B) > P(A)$.
Plot D: $P(A|B) > P(A)$.

4.5 Bayes' Theorem

14. The tree diagram for this scenario is illustrated in Figure 4.4.

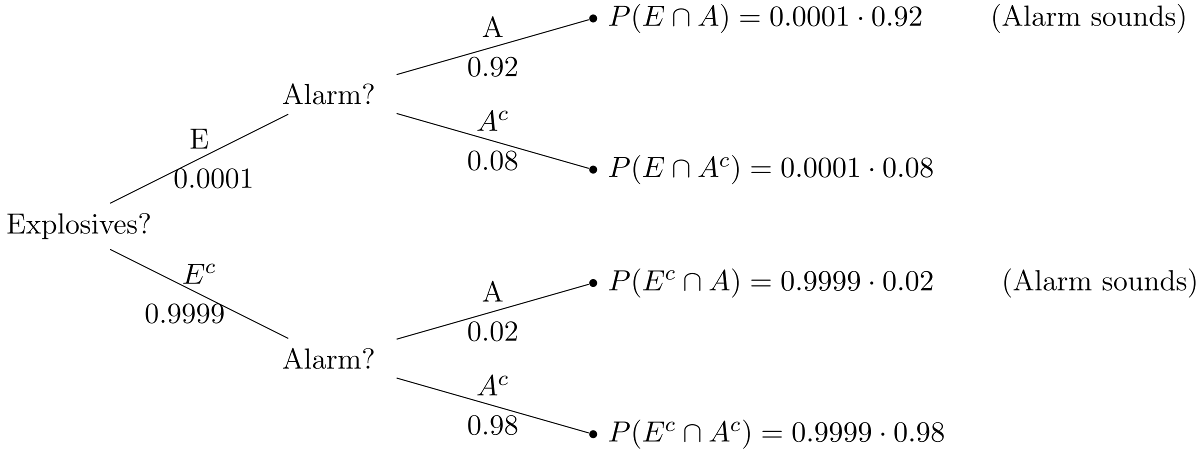


Figure 4.4: The tree diagram for Question 14. E represents the event that the person is carrying explosives and A represents the event that the alarm sounds.

(a) This is the sum of the probabilities of the two branches of the tree diagram that lead to the alarm sounding. More formally:

$$\begin{aligned} P(A) &= P(A \cap E) + P(A \cap E^c) \\ &= P(E)P(A|E) + P(E^c)P(A|E^c) \\ &= 0.0001 \cdot 0.92 + 0.9999 \cdot 0.02 \\ &= 0.02009 \end{aligned}$$

(b) We are given that the alarm has sounded, so either the first branch or third branch of the tree diagram has occurred. What is the conditional probability the first branch has occurred? This is the probability of the first branch, divided by the sum of the probabilities of the first and third branches. More formally:

$$\begin{aligned} P(E|A) &= \frac{P(E \cap A)}{P(A)} \\ &= \frac{P(E)P(A|E)}{P(E)P(A|E) + P(E^c)P(A|E^c)} \\ &= \frac{0.0001 \cdot 0.92}{0.0001 \cdot 0.92 + 0.9999 \cdot 0.02} \\ &= 0.00458 \end{aligned}$$

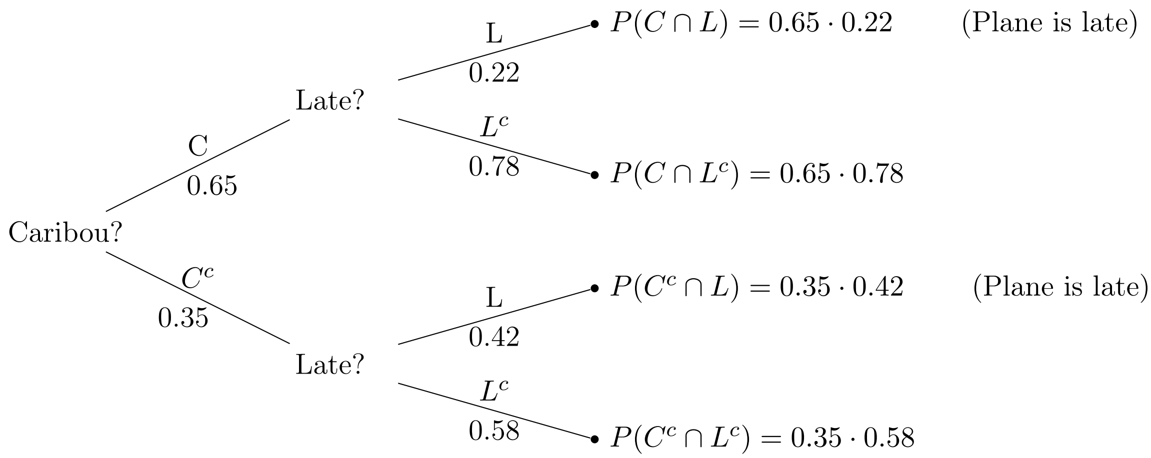


Figure 4.5: The tree diagram for Question 15. C represents the event that the flight is a Caribou Air flight, and L represents the event that the flight is late.

15. The tree diagram for this scenario is illustrated in Figure 4.5.

    (a) This is the sum of the probabilities of the two branches of the tree diagram that lead to a late flight. More formally:

$$\begin{aligned} P(L) &= P(L \cap C) + P(L \cap C^c) \\ &= P(C)P(L|C) + P(C^c)P(L|C^c) \\ &= 0.65 \cdot 0.22 + 0.35 \cdot 0.42 \\ &= 0.29 \end{aligned}$$

    (b) We are given that the flight is late, so either the first branch or third branch of the tree diagram has occurred. What is the conditional probability that the first branch has occurred? This is the probability of the first branch, divided by the sum of the probabilities of the first and third branches. More formally:

$$\begin{aligned} P(C|L) &= \frac{P(C \cap L)}{P(L)} \\ &= \frac{P(C)P(L|C)}{P(C)P(L|C) + P(C^c)P(L|C^c)} \\ &= \frac{0.65 \cdot 0.22}{0.65 \cdot 0.22 + 0.35 \cdot 0.42} \\ &= 0.493 \end{aligned}$$

4.6 Counting rules: Permutations and Combinations

16. (a) $\binom{52}{2} = 1326$. (The order in which the cards are received is unimportant, so this is a combinations problem.)
    (b) $\binom{52}{5} = 2,598,960$. (The order in which the cards are received is unimportant, so this is a combinations problem.)
    (c) As given in problem 16b, the total number of possible hands is $\binom{52}{5} = 2,598,960$. For any given suit, there are $\binom{13}{5} = 1287$ possible flushes. There are 4 suits, so there are $4 \times \binom{13}{5} = 5148$ possible flush hands. Since each of the 2598960 five card hands are equally likely, the probability of getting dealt a flush is $\frac{5148}{2598960} \approx 0.00198$. (About once every 505 deals.)

17. (a) $\binom{200}{4} = 64,684,950$. In a simple random sample, the order of selection is not important (we care only about which people have been selected).
    (b) In a simple random sample every individual has the same chance of being selected, so the probability any individual students gets chosen is $\frac{4}{200} = \frac{1}{50}$. (We could also have used an argument based on combinations. There are $\binom{200}{4}$ possible samples, and $\binom{1}{1} \times \binom{199}{3}$ of these contain a specific student. So the probability any individual student gets selected is $\frac{\binom{1}{1}\binom{199}{3}}{\binom{200}{4}} = \frac{1}{50}$.)

18. (a) $\binom{25}{5} = 53,130$. Here the order of selection is not important (we care only which 5 people have been picked for the committee).
    (b) $P_7^{30} = 10,260,432,000$. Here the order is important (the drivers are being assigned to 7 different routes, so among each group of 7 drivers, there are 7! ways of assigning them to the routes).

4.7 Probability and the Long Run

19. Only A is true.

4.8 Chapter Exercises

4.8.1 Basic Calculations

20. This scenario is illustrated in Figure 4.6.

    (a) $P(A \cup B) = 0.3 + 0.8 - 0.125 = 0.975$.

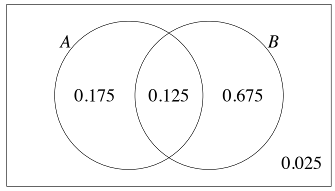


Figure 4.6: Venn diagram for Question 20.

(b) $P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.975 = 0.025$.

(c) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.125}{0.80} = 0.15625$.

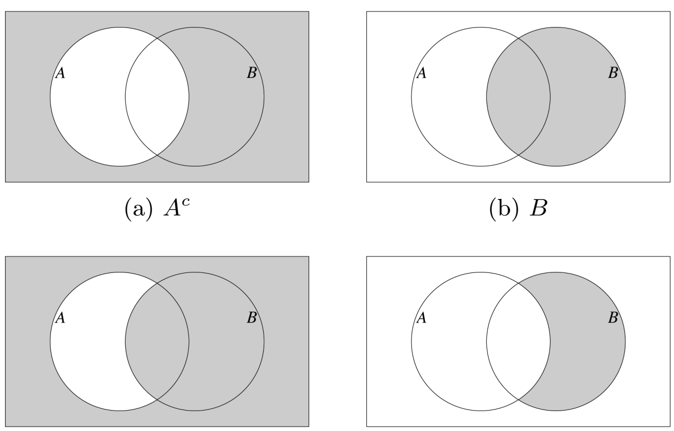
(d) $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.125}{0.30} = 0.4166667$.

(e) No, they are not independent, since $P(A) \neq P(A|B)$.

(f) $P(A^c \cup B) = 0.825$. $A^c \cup B$ is illustrated in Figure 4.7c. From the Venn diagram in Figure 4.6, we can see that the probability of this region is $0.125 + 0.675 + 0.025 = 0.825$.
Alternatively, we could have used the addition rule:

$$\begin{aligned} P(A^c \cup B) &= P(A^c) + P(B) - P(A^c \cap B) \\ &= 0.7 + 0.80 - 0.675 \\ &= 0.825 \end{aligned}$$

(Why is $P(A^c \cap B) = 0.675$? See Figure 4.7d and Figure 4.6. $B = (A^c \cap B) \cup (A \cap B)$, and $P(B) = P(A^c \cap B) + P(A \cap B)$, so $P(A^c \cap B) = 0.8 - 0.125 = 0.675$.)



(a) A^c (b) B

(c) $A^c \cup B$ (d) $A^c \cap B$

Figure 4.7: $A^c \cup B$ and $A^c \cap B$.

21. (a) $P(A \cap B) = 0.02$. By the addition rule, $0.68 = 0.5 + 0.2 - P(A \cap B)$, and this implies $P(A \cap B) = 0.02$. This scenario is illustrated in Figure 4.8.

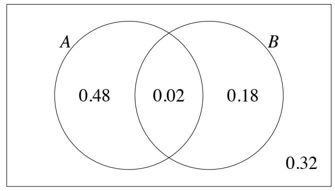


Figure 4.8: Venn diagram for Question 21.

(b) $P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.68 = 0.32$.

(c) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.02}{0.20} = 0.10$.

(d) $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.02}{0.50} = 0.04$.

(e) No, they are not, since $P(A) \neq P(A|B)$.

(f) $P(A^c \cup B^c) = 0.98$. There are different approaches one could use here. One method:

$$P(A^c \cup B^c) = 1 - P(A \cap B) = 1 - 0.02 = 0.98$$

(The only event that is not in $A^c \cup B^c$ is $A \cap B$. See Figure 4.12e for an illustration of this.)

(g) $P(A^c \cap B^c) = 0.32$. There are different approaches one could use here. One method:

$$P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.68 = 0.32$$

($A^c \cap B^c$ is everything that is outside of $A \cup B$. See Figure 4.11e for an illustration of this..)

22. (a) A and B are mutually exclusive. *False, since* $P(A \cap B) = 0.2 \neq 0$.

(b) A and B are independent. *False, since* $P(A \cap B) \neq P(A)P(B)$ $(0.2 \neq 0.2 \times 0.2)$.

(c) $P(A|B) = 0$. *False.* $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.2} = 1$.

(d) $P(A|B) = 1$. *True.* $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.2} = 1$.

(e) A and B^c are not independent. *True.* $P(A|B^c) = 0 \neq P(A)$. *It might help to remember that* A^c *and* B^c *are independent if and only if* A *and* B *are independent.*

(f) $P(A \cup B) = P(A \cap B)$. *True.* $P(A \cup B) = P(A \cap B) = 0.2$.

(g) $P(A|B) = P(B|A)$. *True.* $P(A|B) = P(B|A) = 1$.

23. (a) $\frac{71+19}{100} = 0.9$.

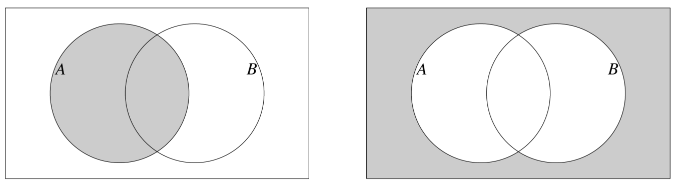
(b) $\frac{71}{73}$.

(c) No, they are not independent. $P(A) = \frac{90}{100}, P(B) = \frac{73}{100}, P(A \cap B) = \frac{71}{100} \neq P(A) \times P(B)$.

24. (a) $P(A|(B \cup A)^c) = 0$, since $P(A \cap (B \cup A)^c) = 0$ (see Figure 4.9).

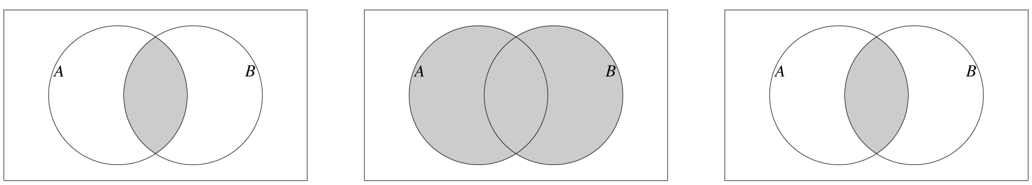
(b)

$$
\begin{aligned}
P(A \cap B | A \cup B) &= \frac{P((A \cap B) \cap (A \cup B))}{P(A \cup B)} \\
&= \frac{P(A \cap B)}{P(A \cup B)} \qquad \text{(See Figure 4.10)} \\
&= \frac{0.2}{0.4 + 0.7 - 0.2} \\
&= \frac{2}{9}
\end{aligned}
$$



(a) A (b) $(B \cup A)^c$

Figure 4.9: Venn diagrams illustrating that A shares no outcomes with $(B \cup A)^c$.



(a) $A \cap B$ (b) $A \cup B$ (c) $(A \cap B) \cap (A \cup B)$

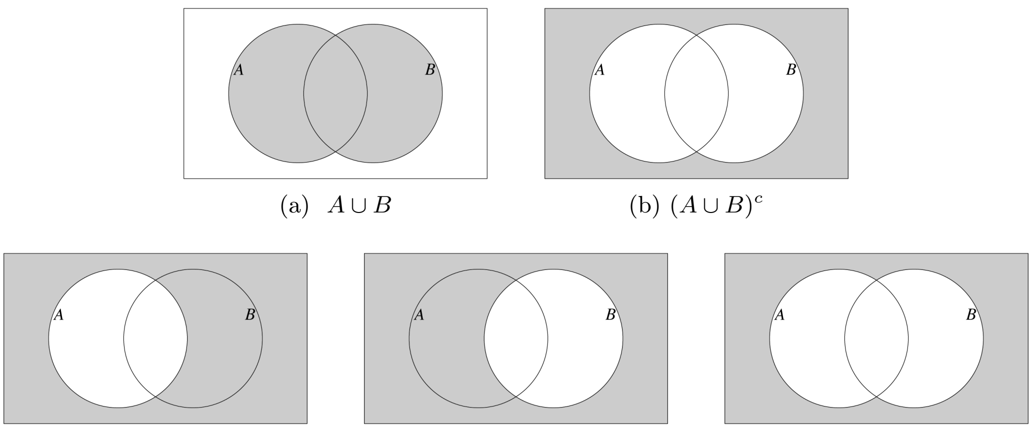
Figure 4.10: Venn diagrams illustrating that $(A \cap B) \cap (A \cup B) = A \cap B$.

25. (a) $\frac{3}{6} = \frac{1}{2}$.
    (b) $\frac{1}{6}$.
    (c) 0. If we are given the number is even, it cannot possibly be a 3.
    (d) $\frac{2}{3}$. If we are given the number is even, we know it is a 2, 4, or 6. Two of these 3 equally likely possibilities are at least 4 (4 and 6), and so $P(\text{The number is at least 4}|\text{The number is a 2}$ $\frac{2}{3}$.
    (e) No. $P(\text{Even}) = \frac{3}{6}$. $P(\text{More than 3}) = \frac{3}{6}$. $P(\text{Even} \cap \text{More than 3}) = \frac{2}{6}$. (The numbers 4 and 6 are more than 3 and even.) $\frac{2}{6} \neq \frac{3}{6} \times \frac{3}{6}$, so these events are not independent.
    (f) 0. If we know the number is at least 4, it cannot possibly be 2 or less. $P(\text{2 or less}|\text{At least 4} = \frac{P(\text{2 or less} \cap \text{at least 4})}{P(\text{at least 4})}) = \frac{0}{3/6} = 0$.

4.8.2 Concepts

26. (a) The odds in favour of rolling a 5 or a 6 are 1:2. (Reduced from 2:4 – odds are usually expressed using the smallest whole numbers that have the appropriate ratio.)
    The odds against rolling a 5 or a 6 are 2:1.
    (b) The odds in favour of drawing a jack are 1:12. The odds against drawing a jack are 12:1.
    (c) $\frac{5}{5+3} = \frac{5}{8} = 0.625$.
    (d) $\frac{1}{19+1} = \frac{1}{20} = 0.05$.

27. (a) See Figure 4.11.
    (b) See Figure 4.12.



(a) $A \cup B$ (b) $(A \cup B)^c$

(c) A^c (d) B^c (e) $A^c \cap B^c$

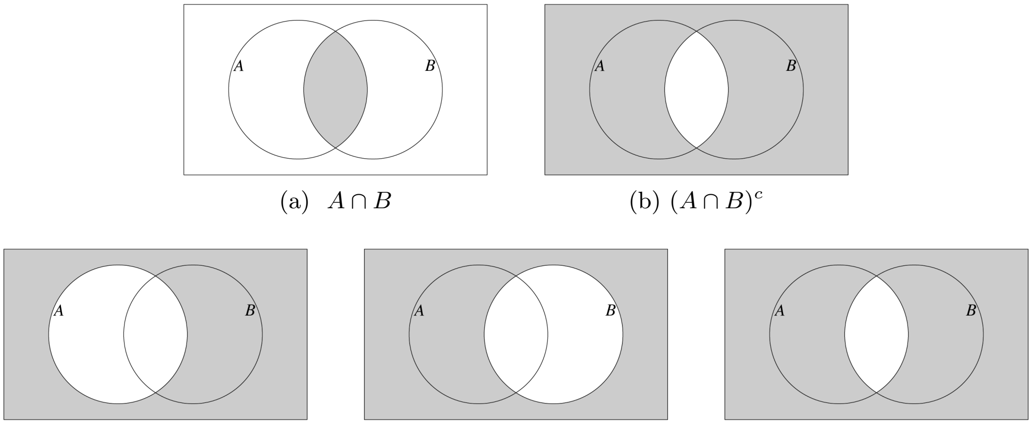
Figure 4.11: Venn diagrams illustrating that $(A \cup B)^c = A^c \cap B^c$.

28.

$$P(A \cup B \cup C) = P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$

29. C (*independent events are never mutually exclusive*). If two events A and B are mutually exclusive, then $P(A \cap B) = 0$. If A and B are independent, $P(A \cap B) = P(A)P(B)$. $P(A)P(B)$ cannot be 0 unless one or more of the events have probability 0, so for mutually exclusive events $P(A \cap B) \neq P(A)P(B)$.

30. (a) If $P(A) = 0.30$ and $P(B) = 0.30$, then $P(A \cap B) = 0.09$. *False.* $P(A \cap B)$ *would equal 0.09 if* A *and* B *are independent, but this statement is not true in general.*



(a) $A \cap B$ (b) $(A \cap B)^c$ (c) A^c (d) B^c (e) $A^c \cup B^c$

Figure 4.12: Venn diagrams illustrating that $(A \cap B)^c = A^c \cup B^c$.

(b) If $P(A) = 0.30$ and $P(B) = 0.30$, and $P(A \cup B) = 0.60$, then A and B are independent. *False. The given probabilities imply $P(A \cap B) = 0$, which does not equal $P(A) \times P(B)$, and thus A and B are not independent.*

(c) If $P(A) = 0.50$, and the probability that B occurs given A occurs is 0.50, then $P(A \cap B) = 0.25$. *True.* $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)}{0.5} = 0.5$. *This implies* $P(A \cap B) = 0.5 \times 0.5 = 0.25$.

(d) If $P(A) = 0$, then A is independent of any event. *True. If $P(A) = 0$, then $P(A|anything) = 0$.*

(e) If $P(B|A) = 0.80$, then $P(B) > 0.80$. *False. The information $P(B|A) = 0.80$ implies only that $0 < P(A \cap B) < 1$ and $0 < P(B) < 1$.*

31. (a) If $P(A) = 0.6$, $P(B) = 0.7$, then $0.3 \leq P(A \cap B) \leq 0.6$. *True. $P(A \cap B)$ cannot be greater than the lesser of $P(A)$ and $P(B)$ (0.6 in this case). Also, $P(A \cup B) \leq 1 \implies P(A) + P(B) - P(A \cap B) \leq 1 \implies P(A \cap B) \geq P(A) + P(B) - 1$. Here, $P(A \cap B) \geq 0.3$. Putting these inequalities together, $0.3 \leq P(A \cap B) \leq 0.6$. (Drawing a Venn diagram can help to visualize this.)*

(b) If $P(A) = 0.5$, then for any event B: $0 \leq P(A \cap B) \leq 0.5$. *True. With no information about B we can say only that $P(A \cap B) \geq 0$ (since it's a probability) and $P(A \cap B) \leq 0.5$ (since the probability of the intersection cannot possibly be greater than $P(A)$.*

(c) If $P(A|B) = 0$, then A and B are mutually exclusive. *True.* $P(A|B) = 0 \implies P(A \cap B) = 0 \implies$ *A and B are mutually exclusive.*

(d) Suppose that $P(A) = 0.99$. Event A and its complement are mutually exclusive and independent. *False. A and A^c are mutually exclusive, but are not independent. $P(A|A^c) = 0 \neq P(A)$.*

(e) If $P(A|B) \neq P(B)$ then A and B are not independent. *False. Knowing*

$P(A|B) \neq P(B)$ gives us very little information about A and B or their probabilities of occurring. Given only the information $P(A|B) \neq P(B)$, we cannot possibly tell if A and B are independent, .

32. No, they are not independent. They are mutually exclusive, so $P(A \cap B) = 0$, but $P(A) \times P(B) = 0.5 \times 0.3 = 0.15 \neq 0$.

33. (a) A and B are independent. (Knowing the outcome of the first toss gives us no information about the outcome of the second toss.)
    (b) A and B^c are independent. (Knowing the outcome of the first toss gives us no information about the outcome of the second toss.)
    (c) A and A^c are not independent. $P(A|A^c) = 0 \neq P(A)$. If we know the coin came up tails (A^c), we know with certainty it did not come up heads (A) on the same toss.

34. A and B are independent in plots B and D.

35. (a) $P(A|B) < P(A)$. *False. Knowing the person weighs more than 100 kg makes if more likely they are male: $P(A|B) > P(A)$. (The proportion of males that weigh more than 100 kg is greater than that of females.)*
    (b) $P(B|A) < P(B)$. *False. Knowing the person is a male makes it more likely they weigh more than 100 kg: $P(B|A) > P(B)$. (The proportion of males that weigh more than 100 kg is greater than that of females.)*
    (c) A and B are mutually exclusive. *False. It is possible for a Canadian adult to be both male and weight more than 100 kg.*
    (d) A and B are not independent. *True, these are dependent events. ($P(B|A) > P(B)$). For them to be independent events, the proportion of men that weigh more than 100 kg would have to be exactly the same as the proportion of women that weigh more than 100 kg.*

36. (a) $P(B|A) > P(B)$. A and B are not independent. Weather systems can be large, and take several days to pass through. If it rains next Tuesday, it is more likely to rain on the next the day than if there were bright sunshine on Tuesday.
    (b) $P(B|A) = P(B)$. These events are independent. Knowing the weather next Tuesday will tell us nothing about the weather 10 years from now.
    (c) $P(B|A) > P(B)$. These events are not independent. People who do not finish high school are more likely to go to prison.

37. (a) No, they are not mutually exclusive. If the first coin comes up heads and the second tails, they will both pass the part. $P(A \cap B) = \frac{1}{4}$.
    (b) Yes, they are independent. $P(A) = \frac{1}{2}, P(B) = \frac{1}{2}$, and $P(A \cap B) = \frac{1}{4}$.

38. (a) The probability that Daniel passes the assignment is a tiny bit greater than

0.40. He has a near zero chance of passing if he is guessing randomly, so his probability of passing is just a bit larger than $0.5 \times 0.8 + 0.5 \times 0 = 0.40$. (Daniel's probability of passing when guessing randomly is 0.0025. We don't yet know how to find this probability, but it should be clear that his probability of passing when guessing randomly is very small.)

(b) No, they are not mutually exclusive. (It is possible that they both pass the test.)

(c) No, they are not independent. If Magnus passes the assignment, Daniel will also pass if he flipped heads. (And almost surely fail if he flipped tails, since he is randomly guessing in that event.) If Magnus fails the assignment, Daniel will almost surely fail. The information that Magnus passes makes it much more likely that Daniel passes. $P(\text{Daniel passes}|\text{Magnus passes}) \approx 0.5$, $P(\text{Daniel passes}|\text{Magnus fails}) \approx 0$.

39. (a) Probabilities can be negative. *False. Probabilities must lie between 0 and 1.*

(b) For any two events A and B, $P(A|B) > P(A)$. *False.* $P(A|B)$ *can be greater than, less than, or equal to* $P(A)$.

(c) If A and B are independent, then $P(A|B) = 0$. *False.*

(d) If two cards are drawn from a standard deck, the probability both cards are red is greater if the cards are drawn *with replacement* than if they were drawn *without replacement. True.* $\frac{26}{52} \times \frac{26}{52} > \frac{26}{52} \times \frac{25}{51}$.

(e) If two events A and B are independent, then their complements (A^c and B^c) are also independent. *True. This isn't terribly difficult to prove, but it requires a step that might more confusing than it's worth. It may help to think of it in these terms: Two events A and B are independent if knowing whether A has occurred or A has not occurred does not change the probability of B. The statement "A has occurred or A has not occurred" is equivalent to the statement "A^c has not occurred or A^c has occurred." If A and B are independent, so are their complements.*

(f) If A and B are mutually exclusive, then $P(A|B) = 0$. *True. If A and B are mutually exclusive,* $P(A \cap B) = 0$, *which implies* $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{0}{P(B)} = 0$.

40. (a) If $P(A) = P(B)$, then A and B are independent. *False. We cannot possibly tell if A and B are independent based solely on the information that* $P(A) = P(B)$.

(b) If $P(A) > P(B)$, then $P(A|B) \geq P(B|A)$. *True.* $P(A|B) = \frac{P(A\cap B)}{P(B)}$ *and* $P(B|A) = \frac{P(A\cap B)}{P(A)}$. *If* $P(A) > P(B)$, *then* $\frac{P(A\cap B)}{P(B)} > \frac{P(A\cap B)}{P(A)}$.

(c) If A and B are independent, $P(A \cup B) = P(A) + P(B)$. *False.*

(d) If A and B are mutually exclusive, $P(A \cup B) = P(A) + P(B)$. *True. If A and B are mutually exclusive,* $P(A \cap B) = 0$ *and so* $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B)$.

(e) If $P(A|B) = 1$, then $P(B|A) = 1$. *False. Consider this counterexample: Roll a* *single die once and let A represent rolling an even number, and let B represent rolling a 6. Here, $P(A|B) = 1$ but $P(B|A) = \frac{1}{3}$. (If B is a subset of A, this does not imply A is a subset of B.)*

4.8.3 Applications

41. (a) $P(\text{At least 1 is mislabelled}) = 1 - P(\text{None are mislabelled}) = 1 - 0.75^4 \approx 0.68$.
    (b) No, it's not a reasonable assumption (for a variety of reasons). For example, some restaurants would be more likely than others to serve mislabelled sushi, and thus mislabelled sushi would likely be clumped together. If, say, the first 3 pieces you get are mislabelled, there is a very good chance the fourth one will also be mislabelled.

42. Let M represent the event the student is male, and J represent the event the student has a job. A Venn diagram illustrating the counts in this situation is given in Figure 4.13.

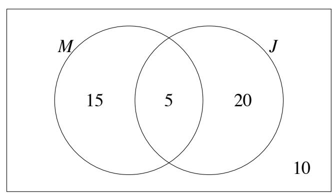


Figure 4.13: Venn diagram for Question 42.

    (a) $P(J^c) = \frac{25}{50} = 0.5$.
    (b) $P(M \cap J^c) = \frac{15}{50} = 0.3$.
    (c) $P(M|J) = \frac{5}{25} = 0.2$.
    (d) $P(M|M^c \cap J) = 0$.
    (e) No, in this question job status and gender are not independent. $P(J) = 0.5$ but $P(J|M) = \frac{5}{20}$. Since $\frac{5}{20} \neq 0.5$, job status is not independent of gender (if a student is randomly selected from this group of students).

43. (a) $\frac{41+27+36+36+32+16}{459} \approx 0.4096$.
    (b) $\frac{133}{459} \approx 0.2898$.
    (c) $\frac{36+16}{41+27+36+36+32+16} \approx 0.2766$.
    (d) $\frac{27+32+36+16}{41+27+36+36+32+16} \approx 0.5904$.
    (e) $\frac{56}{56+37+53} \approx 0.3836$.
    (f) $\frac{27+32}{37+43+27+32} \approx 0.4245$.
    (g) $\frac{27+32+36+16}{37+43+27+32+53+28+36+16} \approx 0.4081$.

(h) No, these events are not independent. $P(\text{University}) = \frac{133}{459} \approx 0.2898$, which differs from $P(\text{University}|\text{Smoker}) = \frac{36+16}{41+27+36+36+32+16} \approx 0.2766$.

(i) No, these events are not mutually exclusive. 52 of the people in the sample have a university education and are smokers.

44. (a) $P(\text{Large}) = \frac{42+507}{6272} = \frac{549}{6272} \approx 0.088$.

(b) $P(\text{Small, moderate, or large}) = \frac{201+2420+209+2769+42+507}{6272} = \frac{6148}{6272} \approx 0.980$.

(c) $P(\text{Cancer}) = \frac{14+201+209+42}{6272} = \frac{466}{6272} \approx 0.074$.

(d) $P(\text{Cancer}|\text{Never/Seldom}) = \frac{14}{14+110} = \frac{14}{124} \approx 0.113$.

(e) $P(\text{Cancer}|\text{Large}) = \frac{42}{42+507} = \frac{42}{549} \approx 0.077$.

(f) $P(\text{Large}|\text{Cancer}) = \frac{42}{14+201+209+42} = \frac{42}{466} \approx 0.090$.

(g) $P(\text{Cancer}|\text{Small or Moderate}) = \frac{201+209}{201+209+2420+2769} = \frac{410}{5599} \approx 0.073$.

45. Let G represent the event that the randomly selected student attended the University of Guelph, and F represent the event that the randomly selected student is female.

(a) $P(G) = \dfrac{16,783}{52,217} = 0.3214$.

(b) $P(F) = \dfrac{27,278}{52,217} = 0.5224$.

(c) $P(G|F) = \frac{10,417}{27,278} = 0.3819$. (This is the number of female U of G students, over the total number of female students.)

(d) $P(F|G) = \frac{10,417}{16,783} = 0.6207$. (This is the number of female University of Guelph students, over the total number of U of G students.)

(e) We require: $P(F|G \cup L^c) = \dfrac{P(F \cap (G \cup L^c))}{P(G \cup L^c)}$

$P(G \cup L^c)$ is the probability that a randomly selected student attended the University of Guelph and not Laurier (which means they attended Guelph or Waterloo):

$$P(G \cup L^c) = \frac{16,783 + 24,377}{52,217} = \frac{41,160}{52,217}$$

$P(F \cap (G \cup L^c))$ is the probability that a randomly selected student is a *female* student at Guelph or Waterloo (not Laurier):

$$P(F \cap (G \cup L^c)) = \frac{10,417 + 10,607}{52,217} = \frac{21,024}{52,217}$$

We can now calculate the final answer:

$$P(F|(G \cup L^c)) = \frac{P(F \cap (G \cup L^c))}{P(G \cup L^c)} = \frac{21,024/52,217}{41,160/52,217} = \frac{21,024}{41,160} = 0.5108$$

46. (a) $P(\text{Good}|\text{Man}) = \frac{23}{30} \approx 0.767$.

(b) $P(\text{Man}|\text{Woman}) = 0$.
(c) No, these events are not independent. $P(\text{Good}) = \frac{40}{65} \approx 0.615$, but $P(\text{Good}|\text{Man}) = \frac{23}{30} \approx 0.767$. (Males were more likely than females to call the dining experience good.)
(d) These events are mutually exclusive and not independent.

47. (a) $0.99 \times 0.98 = 0.9702$.
(b) $(1 - 0.99) \times (1 - 0.98) = 0.0002$.
(c) $P(\text{At least 1 goes off}) = 1 - P(\text{Neither goes off}) = 1 - 0.0002 = 0.9998$.
(d) B or C, as we are interested in the probability of being woken up. (We want the probability of the event that at least one alarm goes off, or the probability of its complement, the event that neither alarm goes off.)

48. (a) $\frac{4}{52} \times \frac{4}{52} = \frac{1}{169} \approx 0.006$.
(b)

$$\begin{aligned} P(\text{Two kings} \mid \text{At least 1 king}) &= \frac{P(\text{Two kings} \cap \text{At least 1 king})}{P(\text{At least 1 king})} \\ &= \frac{P(\text{Two kings})}{1 - P(\text{No kings})} \\ &= \frac{\frac{4}{52} \times \frac{4}{52}}{1 - (\frac{48}{52})^2} \\ &= 0.04 \end{aligned}$$

(c)

$$\begin{aligned} P(\text{Exactly one is red}) &= P(\text{First card is red and second is black}) + P(\text{First card is black and second} \\ &= \frac{26}{52} \times \frac{26}{52} + \frac{26}{52} \times \frac{26}{52} \\ &= 0.5 \end{aligned}$$

(d)

$$\begin{aligned} P(\text{At least one card is red}) &= 1 - P(\text{Neither card is red}) \\ &= 1 - \frac{26}{52} \times \frac{26}{52} \\ &= 0.75 \end{aligned}$$

(e) Drawing a king is independent of drawing a red card (there are 2 red kings and 2 black kings in the deck). The two draws are independent, since the cards are drawn *with* replacement. So:

$$\begin{aligned} P(\text{At least one king and at least one red card}) &= P(\text{At least one king}) \times P(\text{At least one red card}) \\ &= (1 - P(\text{no kings})) \times (1 - P(\text{no red cards})) \\ &= (1 - (\frac{48}{52})^2) \times (1 - (\frac{26}{52})^2) \\ &\approx 0.1109 \end{aligned}$$

49. (a) $\frac{4}{52} \times \frac{3}{51} = \frac{1}{221} \approx 0.0045$.

(b)

$$\begin{aligned} P(\text{Two kings} \mid \text{At least 1 king}) &= \frac{P(\text{Two kings } \cap \text{ At least 1 king})}{P(\text{At least 1 king})} \\ &= \frac{P(\text{Two kings})}{1 - P(\text{No kings})} \\ &= \frac{\frac{4}{52} \times \frac{3}{51}}{1 - \frac{48}{52} \times \frac{47}{51}} \\ &\approx 0.0303 \end{aligned}$$

(c)

$$\begin{aligned} P(\text{Exactly one is red}) &= P(\text{First card is red and second is black}) + P(\text{First card is black and second} \\ &= \frac{26}{52} \times \frac{26}{51} + \frac{26}{52} \times \frac{26}{51} \\ &\approx 0.5098 \end{aligned}$$

(d)

$$\begin{aligned} P(\text{At least one card is red}) &= 1 - P(\text{Neither card is red}) \\ &= 1 - \frac{26}{52} \times \frac{25}{51} \\ &\approx 0.7549 \end{aligned}$$

(e) (One way to approach this is to draw a tree diagram, with 4 branches for each card: red king, black king, red non-king, black non-king.) The probability of getting at least one king and at least one red card is: $\frac{2}{52} \times \frac{51}{51} + \frac{2}{52} \times \frac{26}{51} + \frac{24}{52} \times \frac{4}{51} + \frac{24}{52} \times \frac{2}{51} = \frac{298}{52 \times 51} \approx 0.1123680$.

50. Let F represent the event that Mateo takes his Ferrari, and A represent the event that he gets into an accident. $P(F) = P(2 \text{ heads}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. The tree diagram for this scenario is illustrated in Figure 4.14.

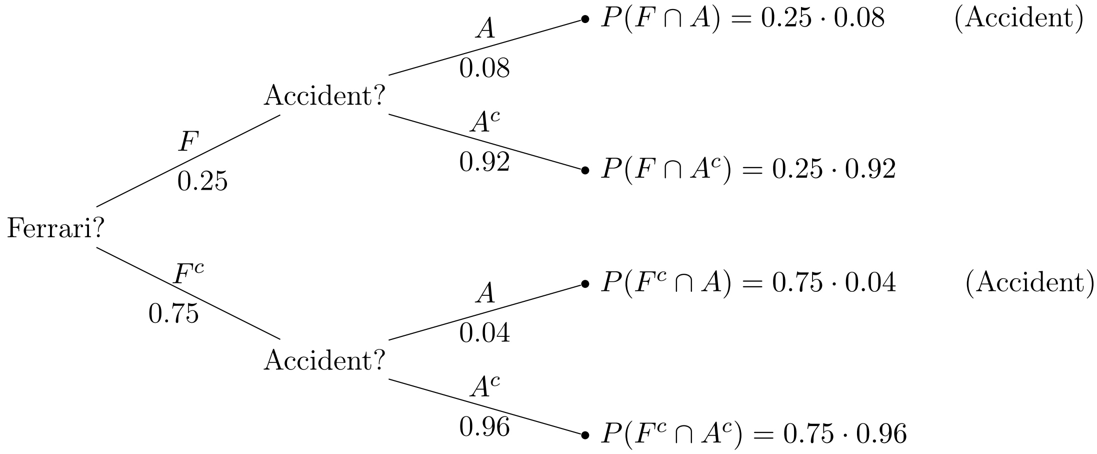


Figure 4.14: The tree diagram for Question 50. F represents the event that Mateo takes his Ferrari and A represents the event that he gets into an accident.

(a) This is the sum of the probabilities of the two branches of the tree diagram that lead to an accident. More formally:

$$\begin{aligned} P(A) &= P(A \cap F) + P(A \cap F^c) \\ &= P(F)P(A|F) + P(F^c)P(A|F^c) \\ &= 0.25 \cdot 0.08 + 0.75 \cdot 0.04 \\ &= 0.05 \end{aligned}$$

(b) 0.08. (This is given in the question.)

(c) We are given that Mateo has gotten into an accident, so either the first branch or third branch of the tree diagram has occurred. What is the conditional probability the first branch has occurred? This is the probability of the first branch, divided by the sum of the probabilities of the first and third branches. More formally:

$$\begin{aligned} P(F|A) &= \frac{P(F \cap A)}{P(A)} \\ &= \frac{P(F)P(A|F)}{P(F)P(A|F) + P(F^c)P(A|F^c)} \\ &= \frac{0.25 \cdot 0.08}{0.25 \cdot 0.08 + 0.75 \cdot 0.04} \\ &= 0.40 \end{aligned}$$

51. (a) $0.8 \cdot 0.02 + 0.05 \cdot 0.04 + 0.15 \cdot 0.14 = 0.039$.

(b) $\frac{0.15 \cdot 0.14}{0.8 \cdot 0.02 + 0.05 \cdot 0.04 + 0.15 \cdot 0.14} \approx 0.538$.

52. (a) $0.55 \cdot 0.40 + 0.35 \cdot 0.90 + 0.10 \cdot 0.55 = 0.59$.
    (b) $\frac{0.55 \cdot 0.40}{0.55 \cdot 0.40 + 0.35 \cdot 0.90 + 0.10 \cdot 0.55} = 0.373$.

53. (a) It often helps to illustrate this type of problem with a tree diagram, as in Figure 4.15.

$$\begin{aligned} P(\text{Respond "Yes"}) &= P(\text{Heads}) + P(\text{Tails} \cap \text{Stolen}) \\ &= 0.5 + 0.5 \cdot 0.2 \\ &= 0.6 \end{aligned}$$

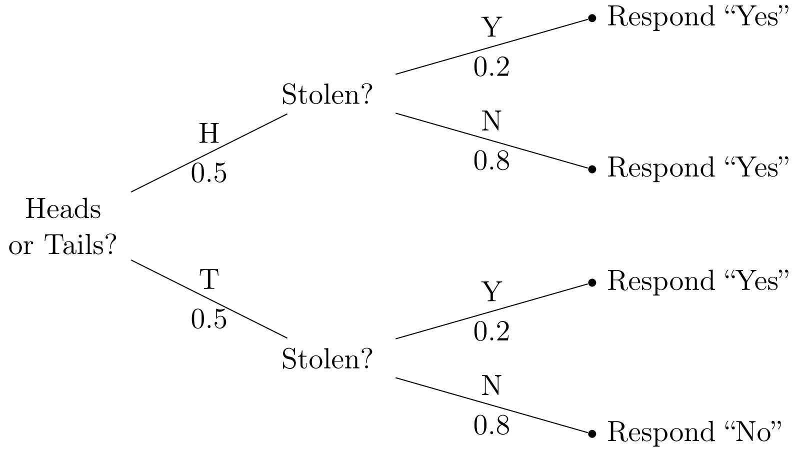


Figure 4.15: Tree diagram for the random response problem.

(b) There are 3 branches of the tree that lead to a "Yes" response. Combined, these 3 branches yield us a probability of "Yes" of 0.60. In two of these 3 branches, the person has actually stolen from their employer.

$$P(\text{Stolen}|\text{Yes}) = \frac{P(\text{Stolen} \cap \text{Yes})}{P(\text{Yes})} = \frac{0.5 \cdot 0.2 + 0.5 \cdot 0.2}{0.60} = \frac{1}{3}$$

This is the ratio of the sum of the probabilities of the branches in which the person has both stolen and responded "Yes", to the probability that a person gives a "Yes" response.

(c) This question is a little bit trickier. We know that:

$$P(\text{Yes}) = P(\text{Heads}) + P(\text{Tails} \cap \text{Stolen}) = P(\text{Heads}) + P(\text{Tails})P(\text{Stolen})$$

($P(\text{Tails} \cap \text{Stolen}) = P(\text{Tails}) \times P(\text{Stolen})$, since getting tails on the coin toss is independent of whether the respondent has actually stolen from their employer.) The probability of getting heads is assumed to be 0.50 (we're assuming a *fair* coin), and thus:

$$P(\text{Yes}) = 0.50 + 0.50 \times P(\text{Stolen})$$

Since the best estimate of $P(\text{Yes})$ is 0.67 (67% of respondents said yes), the best estimate of $P(\text{Stolen})$ is $\frac{0.67-0.5}{0.5} = 0.34$.
Keep in mind that most people find it easier to use the tree diagram approach for these types of questions.

54. The probability of getting a "yes" response is:

$$P(\text{Yes}) = \frac{4}{6} + \frac{2}{6} \times P(\text{Cheat})$$

Since 82% of people said "yes", the estimated probability of a "yes" response is 0.82, and so:

$$0.82 = \frac{4}{6} + \frac{2}{6} \times \text{Estimated proportion that cheated}$$

So the estimated proportion of these people who have cheated is: $\frac{0.82-\frac{4}{6}}{\frac{2}{6}} = 0.46$.

55. (a) $P(\text{none have the disorder}) = (1 - \frac{1}{1000})^{30} = 0.9704$.
(b) You will have to sample more than 500 mice if the first 500 mice do not have the disorder:

$$P(\text{First 500 do not have the disorder}) = 0.999^{500} = 0.6064$$

Note that this probability is quite a bit bigger than 50%. Surprising?
(c) If we sample k mice, the probability that at least one has the disorder is:

$$P(\text{At least one has the disorder}) = 1 - P(\text{None have the disorder}) = 1 - 0.999^k$$

We need to find the minimum value of k such that $1 - 0.999^k \geq 0.50$. This implies we need the value of k such that $0.999^k \leq 0.50$. We could use trial-and-error, or we could find k a little more efficiently using logarithms:

$$0.999^k \leq 0.50 \implies k \cdot \log(0.999) \leq \log(0.50) \implies k \geq \frac{\log(0.50)}{\log(0.999)} \implies k \geq 692.8.$$

We would need to sample at least 693 mice to have at least a 50% chance of finding a mouse with the defect. Would you have guessed a value this large?

56. (a) $P(\text{System failure}) = P(\text{All components fail}) = 0.32^4 \approx 0.01048$.
(b) Here we need to find k such that $0.32^k < 0.0001$. We could use trial and error, but it's usually a little more efficient to use properties of logs: $0.32^k < 0.0001 \implies k ln(0.32) < ln(0.0001) \implies k > \frac{ln(0.0001)}{ln(0.32)} \implies k > 8.083257$. We would need at least 9 components connected in parallel to ensure the probability of failure is less than 0.0001.

57. (a) $0.77^5 = 0.2707$.

(b) $0.77^{10} = 0.0732$.

(c) Here we need to find k such that $1 - 0.77^k \geq 0.99990$. We could use trial and error, but it's usually a little more efficient to use properties of logs: $1 - 0.77^k \geq 0.99990 \implies 0.77^k \leq 0.0001 \implies kln(0.77) \leq ln(0.0001) \implies k \geq \frac{ln(0.0001)}{ln(0.77)} \implies k > 35.23941$. Since k must be an integer, we would need at least 36 components connected in parallel to have a probability of failure this low.

58. (a) $0.9999 \cdot 0.0001 = 0.00009999$.

(b) $0.9999^9 \cdot 0.0001 = 0.0000991$.

(c) This will occur as long as the first 2000 do not all not have the disorder: $P(\text{First 2000 have none with the disorder}) = 0.9999^{2000}$. The probability we are looking for here is the complement of this: $P(\text{First 2000 have at least one with the disorder}) = 1 - 0.9999^{2000} = 0.181$.

59. (a) $\frac{178}{200} = 0.89$.

(b) $\frac{20}{4+20} = 0.833$.

(c) No. $P(\text{A thought it a forgery}) = 0.11$, $P(\text{A thought it a forgery}|\text{B thought it a forgery}) = 0.8333$. Even though they are viewing the piece *independently* (using the English language meaning of the term), they are experts, and knowing that one of them thought the piece a forgery makes it much more likely it *is* a forgery, which makes it more likely the other will think so as well.

60. Let J represent the event that John goes to the party, and S represent the event that Stephanie goes to the party. In the question we are given the conditional probabilities $P(J|S) = 0.94$ and $P(J|S^c) = 0.03$.

(a) This scenario is illustrated in Figure 4.16.
$P(J) = P(S)P(J|S) + P(S^c)P(J|S^c) = 0.8 \cdot 0.94 + 0.2 \cdot 0.03 = 0.758$. (This is the sum of the probabilities of the first and third branches in the tree diagram.)

(b) This scenario is illustrated in Figure 4.17.

$$\begin{aligned} P(J) &= 0.38 \\ \implies P(S)P(J|S) + P(S^c)P(J|S^c) &= 0.38 \\ \implies P(S) \cdot 0.94 + (1 - P(S)) \cdot 0.03 &= 0.38 \\ \implies P(S) &= 0.3846 \end{aligned}$$

(c) They are not mutually exclusive since they can both attend the party. They are not independent; John is *much* more likely to go to the party if Stephanie goes.

61. Most people find it helpful to draw a tree diagram for this type of problem (see Figure 4.18).

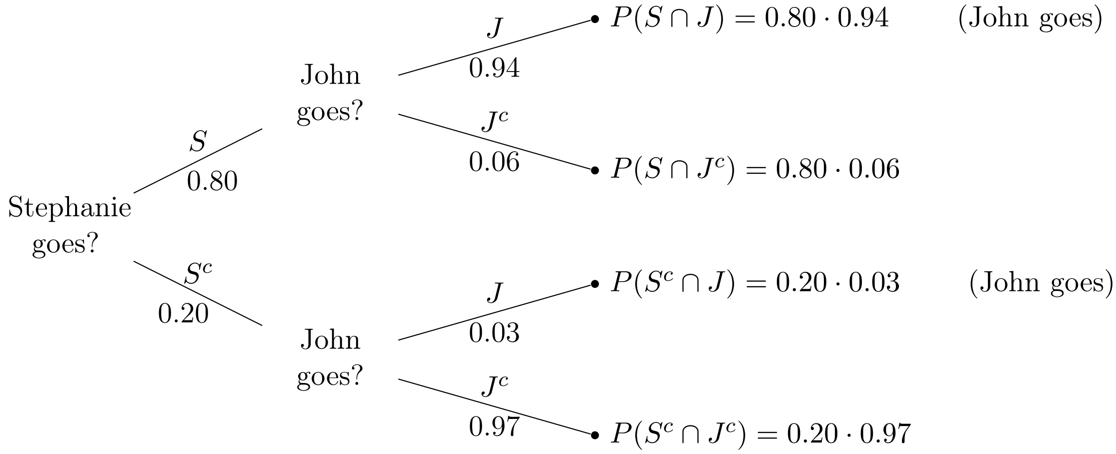


Figure 4.16: The tree diagram for Question 60a.

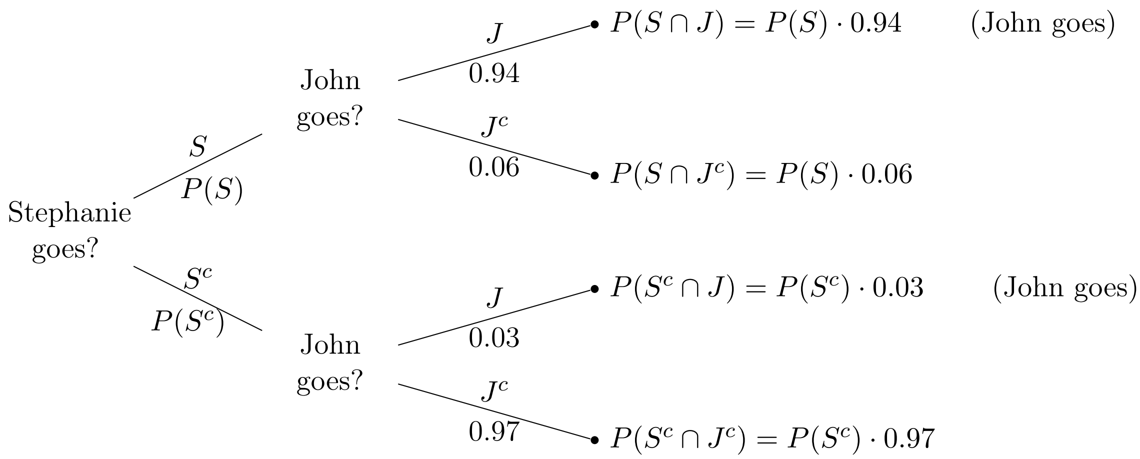


Figure 4.17: The tree diagram for Question 60b.

$$\begin{aligned}
P(+\text{ve test}) &= P(\text{Defect} \cap +\text{ve test}) + P(\text{No defect} \cap +\text{ve test}) \\
&= 0.00001 \cdot 0.99 + (1 - .00001) \cdot 0.005 \\
&= 0.00500985
\end{aligned}$$

(a)

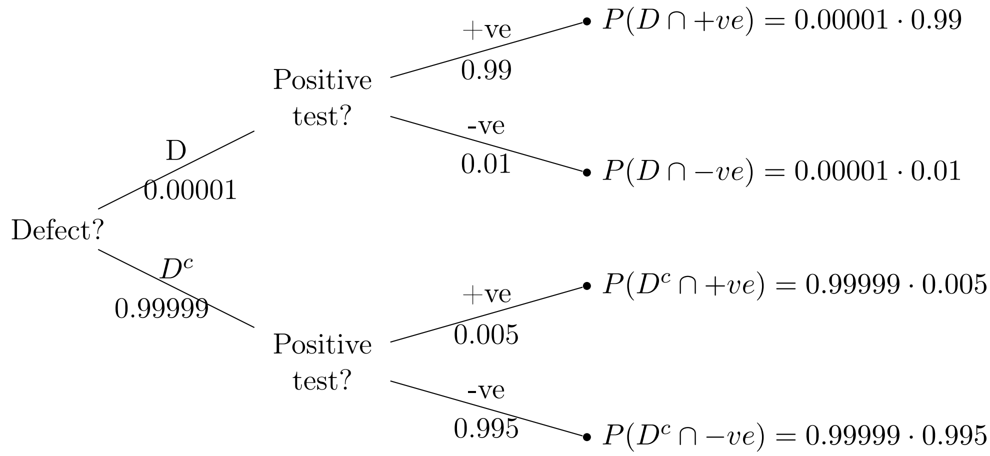


Figure 4.18: Tree diagram for the genetic defect problem.

(b)

$$\begin{aligned} P(\text{Defect} \mid +\text{ve test}) &= \frac{P(\text{Defect} \cap +\text{ve test})}{P(+\text{ve test})} \\ &= \frac{0.00001 \cdot 0.99}{0.00001 \cdot 0.99 + (1 - 0.00001) \cdot 0.005} \\ &= 0.001976 \end{aligned}$$

Note that there is very little chance the person has the defect, even though they tested positive.

62. (a) 0.99726, as given in the table.
    (b) $1 - 0.99726 = 0.00274$.
    (c) This is impossible to determine from the given information (but could be figured out from similar information given at the StatsCan site).
    (d) $0.99775 \cdot 0.99751 \cdot 0.99726 \cdot (1 - 0.99702) = 0.002957765$.
    (e) $0.99775 \cdot 0.99751 \cdot 0.99726 = 0.9925386$.

Chapter 5

Discrete Random Variables and Discrete Probability Distributions

J.B.'s strongly suggested exercises: 2, 5, 6, 7, 8, 10, 13, 14, 15, 19, 20, 22, 23, 25, 26, 28, 30, 48, 51, 53, 54, 57, 58, 62, 64, 66 (a-d).

5.1 Introduction

5.2 Discrete and Continuous Random Variables

1. (a) Discrete.
   (b) Continuous.
   (c) Continuous.
   (d) Discrete.
   (e) Continuous.
   (f) Discrete.
   (g) Discrete.

5.3 Discrete Probability Distributions

2. X can take on the values 0, 1, and 2.
   $P(X = 0) = \frac{39}{52} \times \frac{38}{51} = \frac{1482}{2652} = \frac{19}{34}$. (Don't spend much time trying to reduce the fractions in this question—focus on the topic at hand.)

$P(X=1) = \frac{13}{52} \times \frac{39}{51} + \frac{39}{52} \times \frac{13}{51} = \frac{1014}{2652} = \frac{13}{34}$
$P(X=2) = \frac{13}{52} \times \frac{12}{51} = \frac{156}{2652} = \frac{1}{17}$.
In summary, the probability distribution of X is:

x	0	1	2
$p(x)$	$\frac{19}{34}$	$\frac{13}{34}$	$\frac{1}{17}$

3. (a) Discrete.
   (b) 0.1 (since the probabilities must sum to 1).
   (c) The most likely value is 30.
   (d) $P(X < 32) = P(X = 10) + P(X = 20) + P(X = 30) = 0.2 + 0.2 + 0.5 = 0.9$.
   (e) $P(X < 25|X < 35) = \frac{0.4}{0.9} \approx 0.44$.

$$\begin{aligned} P(X < 25|X < 35) &= \frac{P(X < 25 \cap X < 35)}{P(X < 35)} \\ &= \frac{P(X = 10 \text{ or } X = 20)}{P(X = 10 \text{ or } X = 20 \text{ or } X = 30)} \\ &= \frac{0.2 + 0.2}{0.2 + 0.2 + 0.5} \\ &= \frac{0.4}{0.9} \end{aligned}$$

4. (a) and (c) are valid discrete probability distributions, since for these distributions the probabilities all lie between 0 and 1 and they sum to 1. (Note that some discrete random variables can take on negative values and non-integer values.) (b) is not a valid discrete probability distribution, since the probabilities do not sum to 1. (d) is not a valid discrete probability distribution, since one of the probabilities listed in the table is negative.

5.3.1 The Expectation and Variance of Discrete Random Variables

5.3.1.1 Calculating the expected value and variance of a discrete random variable

5. (a) X is discrete (it takes on only 5 possible values).
   (b) 0.30 (since the probabilities must sum to 1).
   (c) 5.0.
   (d) $P(-1 < X < 4) = P(X = 0) + P(X = 2.5) = 0.1 + 0.1 = 0.2$.
   (e) $E(X) = \sum xp(x) = -2.5 \cdot 0.10 + 0 \cdot 0.10 + 2.5 \cdot 0.10 + 5.0 \cdot 0.40 + 7.5 \cdot 0.30 = 4.25$.

(f) $\sigma = 3.172$.
$\sigma^2 = E[(X-\mu)^2] = \sum(x-\mu)^2 p(x) = (-2.5-4.25)^2 \cdot 0.10 + (0-4.25)^2 \cdot 0.10 + (2.5-4.25)^2 \cdot 0.10 + (5.0-4.25)^2 \cdot 0.40 + (7.5-4.25)^2 \cdot 0.30 = 10.0625$.
$\sigma = \sqrt{10.0625} = 3.172$.

6. Let X represent the number of babies.

(a) $\mu = \sum xp(x) = 1 \cdot 0.988 + 2 \cdot 0.01187 + 3 \cdot 0.000128 + 4 \cdot 0.000002 = 1.012132$.

(b) $\sigma = 0.11069$.
$E(X^2) = \sum x^2 p(x) = 1^2 \cdot 0.988 + 2^2 \cdot 0.01187 + 3^2 \cdot 0.000128 + 4^2 \cdot 0.000002 = 1.036664$. $\sigma^2 = E(X^2) - (E(X))^2 = 1.036664 - 1.012132^2 = 0.01225281$.
$\sigma = \sqrt{0.01225281} = 0.11069$.

(c) $P(X \geq 2) = 0.01187 + 0.000128 + 0.000002 = 0.012$. Equivalently, $P(X \geq 2) = 1 - P(X = 1) = 1 - 0.988 = 0.012$.

(d) $P(X = 2|X \geq 2) = \dfrac{0.01187}{0.01187 + 0.000128 + 0.000002} = 0.989$.

(e) $P(X = 3|X \geq 3) = \dfrac{0.000128}{0.000128 + 0.000002} = 0.985$.

7. (a) Probabilities of impossible events can be negative. *False. (Probabilities must lie between 0 and 1.)*

(b) If a random variable can take on an infinite number of possible values, then it cannot be a discrete random variable. *False. (Some discrete random variables take on a* countably infinite *number of possible values.)*

(c) The mean of a discrete random variable cannot be negative. *False. (Some discrete random variables take on negative values and have negative means.)*

(d) The standard deviation of a discrete random variable cannot be negative. *True. (Standard deviations cannot be negative.)*

(e) The mean of a discrete random variable cannot be greater than its standard deviation. *False. (The mean may be greater than, less than, or equal to the standard deviation.)*

5.3.1.2 Properties of Expectation and Variance

8. (a) $\mu_{X+Y} = \mu_X + \mu_Y = 8.7 + 14.9 = 23.6$.

(b) $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{18.5^2 + 21^2} = 27.9866$.

(c) $\sigma_{X-Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{18.5^2 + 21^2} = 27.9866$.

9. Yes, they are independent. Since the Professor of Course B is simply assigning grades at random, knowing a student's grade in that course yields no information about their grade in course A.

10. Let X represent the time to completion of Step 1 ($\mu_X = 6, \sigma_X = 1.2$).

Let Y represent the time to completion of Step 2 ($\mu_Y = 125, \sigma_Y = 15$).

Let Z represent the time to completion of Step 3 ($\mu_Z = 17, \sigma_Z = 3$).

Let T represent the time to completion of the entire process: $T = X + Y + Z$. We need to find the mean and standard deviation of T.

(a) $E(T) = E(X+Y+Z) = E(X)+E(Y)+E(Z) = 6+125+17 = 148$ minutes.
(b) Since the 3 steps are assumed to be independent, $Var(T) = Var(X+Y+Z) = Var(X) + Var(Y) + Var(Z) = 1.2^2 + 15^2 + 3^2 = 235.44$ minutes^2. $\sigma_T = \sqrt{235.44} = 15.344$ minutes.
(c) Dividing T by 60 changes the scale of measurement to hours. $E(\frac{T}{60}) = \frac{E(T)}{60} = \frac{148}{60} = 2.467$ hours. $Var(\frac{T}{60}) = \frac{Var(T)}{60^2} = 0.0654$ hours^2. $\sigma_{T/60} = \frac{\sigma_T}{60} = \frac{15.344}{60} = 0.2557$ hours.

5.4 The Bernoulli Distribution

11. (a) Let X represent the number of eights that are drawn. Then X has the Bernoulli distribution with $p = \frac{1}{13}$. That is:

$$P(X = x) = (\frac{1}{13})^x(\frac{12}{13})^{1-x} \qquad \text{for } x = 0, 1.$$

(b) $\mu = p = \frac{1}{13}$.
(c) $\sigma = \sqrt{p(1-p)} = \sqrt{\frac{1}{13} \cdot \frac{12}{13}} = \sqrt{\frac{12}{169}} \approx 0.266$.

5.5 The Binomial Distribution

12. (a) $P(X = 3) = \binom{15}{3}0.2^3(1-0.2)^{12} = 0.250$.
(b) Using the binomial probability mass function with $n = 15$ and $p = 0.2$:

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.0352 + 0.1319 + 0.2309 + 0.2501 \\ &= 0.648 \end{aligned}$$

(c) $P(X > 3) = 1 - P(X \leq 3) = 1 - 0.648 = 0.352$.
(d) $P(2 \leq X < 5) = P(X = 2) + P(X = 3) + P(X = 4) = 0.231 + 0.250 + 0.188 = 0.669$.

(e) $P(X = 3|X \leq 4) = \dfrac{P(X = 3 \cap X \leq 4)}{P(X \leq 4)} = \dfrac{P(X = 3)}{P(X \leq 4)}$
$= \dfrac{P(X = 3)}{P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)} = \dfrac{0.2501}{0.8358} =$ 0.2993.

(f) $\mu_X = E(X) = np = 15 \cdot 0.2 = 3$.

(g) $\sigma = \sqrt{np(1-p)} = \sqrt{15 \cdot 0.2 \cdot 0.8} = 1.549$.

13. Let X represent the number of newborns with major structural or genetic birth defects. Then X has a binomial distribution with parameters $n = 50$ and $p = 0.03$.

(a) $P(X = 2) = \binom{50}{2} 0.03^2 (1 - 0.03)^{48} = 0.2555$.

(b)

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{50}{0} 0.03^0 (1-0.03)^{50} + \binom{50}{1} 0.03^1 (1-0.03)^{49} + \binom{50}{2} 0.03^2 (1-0.03)^{48} \\ &= 0.2181 + 0.3372 + 0.2555 \\ &= 0.8108 \end{aligned}$$

(c) $P(X = 10) = \binom{50}{10} 0.03^{10} (1 - 0.03)^{40} = 1.8 \times 10^{-6}$.

(d) $E(X) = np = 50 \cdot 0.03 = 1.5$.

(e) $\sigma = \sqrt{np(1-p)} = \sqrt{50 \cdot 0.03 \cdot 0.97} = 1.206$.

14. Which of the following statements are true?

(a) The variance of a binomial distribution is always greater than the mean. *False. (The variance ($np(1-p)$) must be less than or equal to the mean (np).)*

(b) A binomial random variable can be negative. *False. (A binomial random variable represents a count, and its value must be at least 0.)*

(c) The mean of a binomial random variable can be negative. *False.*

(d) The variance of a binomial random variable can be negative. *False.*

(e) A binomial random variable is a discrete random variable. *True.*

(f) Every discrete random variable is a binomial random variable. *False. (There are many other discrete probability distributions.)*

(g) A binomial random variable represents a count. *True.*

(h) The mean of a binomial random variable is np. *True.*

(i) The standard deviation of a binomial random variable is $\sqrt{np(1-p)}$. *True.*

(j) For any given n, the variance of a binomial random variable is greatest when $p = 0.5$. *True, since $p(1-p)$ is greatest when $p = 0.5$.*

5.5.1 Binomial or Not?

15. (a) Not binomial. The number of hearts would not have a binomial distribution, since the trials are not independent (the probability of getting a heart will change as cards are removed from the deck).
    (b) Binomial. The number of times heads comes up will have a binomial distribution with $n = 10$ and p equal to the probability of heads on any individual toss.
    (c) Not binomial. The number of putts that are holed will not have a binomial distribution. The number of trials is fixed, each trial is either a success or a failure, and we are counting the number of successes, but the probability of success is changing and the trials are not independent.
    (d) Not binomial. The number of newborn babies would not have a binomial distribution. The conditions of the binomial distribution are definitely not satisfied here (there is not even a fixed number of trials).
    (e) Not binomial. The amount of money withdrawn would definitely not have a binomial distribution. (This random variable is not even a count of a number of successes.)

16. (a) Binomial. All of the conditions of a binomial distribution are satisfied here.
    (b) Not binomial. While this random variable is a count (that almost always takes on the value 10), it does not represent the number of successes in a fixed number of trials.
    (c) Binomial. The probability of rolling a sum of 7 on a pair of dice is $\frac{1}{6}$, so here the number of sevens would have a binomial distribution with $n = 14$ and $p = \frac{1}{6}$.
    (d) Not binomial. While the number of children the woman has is a count, it does not represent the number of successes in a fixed number of trials.
    (e) Not binomial. While the number of times the person uses the washroom is a count, it does not represent the number of successes in a fixed number of trials.
    (f) Not binomial. The weight is not a count.

5.5.2 A Binomial Example with Probability Calculations

17. (a) 0.2499. If they are randomly pulling cards, the number of times the king of spades is drawn has a binomial distribution with $n = 50$ and $p = \frac{1}{52}$. $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 0.2499$.
    (b) No. This experiment yields no evidence that they *have* ESP, but we cannot say there is strong evidence they do not. There is, however, strong evidence against their claimed probability of drawing the king of spades. If their claimed

value of $p = 0.2$ is true, the probability they draw the king of spades 2 or fewer times in 50 attempts is only 0.00088.

5.6 The Hypergeometric Distribution

18. Let X represent the number of white balls drawn. Then X has a hypergeometric distribution with $a = 22$, $N = 40$, and $n = 6$.

    (a) $P(X = 4) = \dfrac{\binom{22}{4}\binom{18}{2}}{\binom{40}{6}} = 0.2916.$

    (b) $P(X \leq 1) = P(X = 0) + P(X = 1) = \dfrac{\binom{22}{0}\binom{18}{6}}{\binom{40}{6}} + \dfrac{\binom{22}{1}\binom{18}{5}}{\binom{40}{6}} = 0.00484 + 0.04911 = 0.05395.$

    (c) The mean number of white balls is $6 \cdot \dfrac{22}{40} = 3.3$. The mean number of red balls is $6 \cdot \dfrac{18}{40} = 2.7$.

19. Here we are sampling without replacement (each person can be hired only once). Let X represent the number of hires that have a criminal record. Then X has a hypergeometric distribution with $a = 7$, $N = 20$, and $n = 5$.

    (a) Here we must draw 3 criminals (from 7 total) and 2 non-criminals (from 13 total):
    $P(X = 3) = \dfrac{\binom{7}{3}\binom{13}{2}}{\binom{20}{5}} = 0.1761.$

    (b) Here we must draw 2 criminals (from 7 total) and 3 non-criminals (from 13 total):
    $P(X = 2) = \dfrac{\binom{7}{2}\binom{13}{3}}{\binom{20}{5}} = 0.3874.$

    (c) $P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{\binom{7}{0}\binom{13}{5}}{\binom{20}{5}} = 1 - 0.0830 = 0.9170.$

    (d) $\mu = n \cdot \frac{a}{N} = 5 \cdot \frac{7}{20} = 1.75.$

20. (a) The number of males in the sample has a hypergeometric distribution. The binomial distribution can be used to approximate the hypergeometric, but only if the sample size is a small proportion of the population size (less than 5% or so). Here the sample size is too large (we're drawing 8 from 20), so the binomial distribution would not provide a reasonable approximation. (Using the hypergeometric distribution: $P(X = 4) = 0.3501$. Using the binomial approximation: $P(X = 4) \approx 0.2734$.)

(b) The number of white balls follows a hypergeometric distribution, but since we are drawing only a small proportion of the population (12 from 300), the binomial distribution would provide a reasonable approximation. (Using the hypergeometric distribution: $P(X = 3) = 0.2141$. Using the binomial approximation: $P(X = 4) \approx 0.2120$.)

(c) Here the number of kings has a binomial distribution with $n = 60$ and $p = \frac{1}{13}$, so no approximation is required. $P(X = 7) = 0.088$.

5.7 The Poisson Distribution

5.7.1 Introduction

21. (a) $P(X = 2) = \frac{5^2 e^{-5}}{2!} = 0.0842$.

(b) 0.1247.

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} + \frac{5^2 e^{-5}}{2!} \\ &= 0.0067 + 0.0337 + 0.0842 \\ &= 0.1247 \end{aligned}$$

(c) $P(X > 2) = 1 - P(X \leq 2) = 1 = 0.1247 = 0.8753$.

(d) $P(2 < X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5) = 0.140 + 0.175 + 0.175 = 0.491$.

(e) $P(X = 2 | X \leq 3) = \dfrac{P(X = 2 \cap X \leq 3)}{P(X \leq 3)} = \dfrac{P(X = 2)}{P(X \leq 3)} = \frac{0.08422}{0.2650} = 0.3178$.

(f) $\mu = 5$, since $\mu = \lambda$ for a Poisson random variable.

(g) $\sigma = \sqrt{5}$, since $\sigma = \sqrt{\lambda}$ for a Poisson random variable.

22. (a) $P(X = 1) = \dfrac{4^1 e^{-4}}{1!} = 0.07326$. (The number of shark attacks has a Poisson distribution with $\lambda = 4$.)

(b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.2381$. (Calculated with the Poisson probability mass function with $\lambda = 4$.)

(c) $P(X = 4) = \dfrac{8^4 e^{-8}}{4!} = 0.05725$. (The number of shark attacks in a two-year period has a Poisson distribution with $\lambda = 2 \times 4 = 8$.)

23. Which of the following statements are true?

(a) A Poisson random variable can take on a countably infinite number of possible values. *True. (A Poisson random variable takes on the values 0, 1, 2, ...)*

(b) A Poisson random variable can take on negative values. *False.*
(c) A Poisson random variable represents a count of the number of occurrences of an event. *True.*
(d) The mean and variance of a Poisson random variable are always equal. *True.*
(e) If X has a Poisson distribution, then $P(X = 0) < P(X = 1)$. *False. (Whether or not P(X=0) is less than P(X=1) depends on the value of λ.)*
(f) Every random variable that represents a count is a Poisson random variable. *False.*

5.7.2 The Relationship Between the Poisson and Binomial Distributions

24. Situation (d), where $n = 500$ and $p = 0.01$. The Poisson approximation to the binomial works best when n is large and p is small.

25. (a) Using the binomial pmf: $P(X = 2) = \binom{100}{2} 0.01^2 \cdot 0.99^{98} = 0.18486$.
For the Poisson approximation, $\lambda = np = 100 \cdot 0.01 = 1$ and: $P(X = 2) = \dfrac{1^2 e^{-1}}{2!} = 0.18394$.

(b) Binomial: $P(X \leq 2) = 0.92063$. Poisson approximation: $P(X \leq 2) \approx 0.91970$.
Using the binomial pmf:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{100}{0} 0.01^0 \cdot 0.99^{100} + \binom{100}{1} 0.01^1 \cdot 0.99^{99} + \binom{100}{2} 0.01^2 \cdot 0.99^{98} \\ &= 0.36603 + 0.36973 + 0.18486 \\ &= 0.92063 \end{aligned}$$

Using the Poisson approximation:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{1^0 e^{-1}}{0!} + \frac{1^1 e^{-1}}{1!} + \frac{1^2 e^{-1}}{2!} \\ &= 0.36788 + 0.36788 + 0.18394 \\ &= 0.91970 \end{aligned}$$

(c) Binomial: $\sigma^2 = np(1 - p) = 100 \cdot 0.01 \cdot 0.99 = 0.99$.
Poisson approximation: $\sigma^2 = \lambda = np = 100 \cdot 0.01 = 1$.

5.7.3 Poisson or Not? More Discussion on When a Random Variable has a Poisson distribution

26. The number of students would not follow a Poisson distribution, since students do not arrive randomly and independently. Students are likely to travel in groups, and thus there would be "clumping" of students. It is likely that there would also be time effects (for example, the theoretical rate at which students are arriving might be very different at 12:01 and 12:29).

27. The number of cows in a randomly selected 100 m^2 area of the pasture would *not* follow a Poisson distribution, since the cows would not be distributed randomly and independently (cows tend to stick together).

5.8 The Geometric Distribution

28. Let X be the number of donors required to get the first one with type AB blood.

    (a) $P(X = 5) = (1 - 0.03)^{5-1} \cdot 0.03 = 0.0266$.

    (b) 0.0873. This answer can be found in 2 ways. One way is brute force method:

    $$\begin{aligned} P(X \leq 3) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.97^0 \cdot 0.03 + 0.97^1 \cdot 0.03 + 0.97^2 \cdot 0.03 \\ &= 0.0873 \end{aligned}$$

    Another way is to use the cumulative distribution function:

    $$\begin{aligned} & P(X \leq x) = 1 - (1 - p)^x \\ \implies & P(X \leq 3) = 1 - (1 - 0.03)^3 \\ & \qquad\qquad\; = 0.0873 \end{aligned}$$

    (c) 0.401. This could be answered using the brute force method: $P(X > 30) = 1 - P(X \leq 30) = 1 - [P(X = 1) + P(X = 2) + \ldots + P(X = 30)]$. However, this would require calculating 30 geometric probabilities which would be very cumbersome without using software. It is easier to answer this question if we recognize that the first success will come after the 30th person *if the first 30 people do not have blood type AB*. So $P(X > 30) = 0.97^{30} = 0.401$.

    (d) $\mu = \frac{1}{p} = \frac{1}{0.03} = 33.\dot{3}$.

    (e) $\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.03}{0.03^2}} = 32.83$.

29. Let X represent the required number of cups to get the first winner.

(a) $P(X = 4) = (\frac{5}{6})^{4-1}(\frac{1}{6}) = 0.0965$.

(b) 0.5177. This answer can be found in 2 ways. One way is brute force method:

$$\begin{aligned} P(X \leq 4) &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\ &= (\frac{5}{6})^0(\frac{1}{6}) + (\frac{5}{6})^1(\frac{1}{6}) + (\frac{5}{6})^2(\frac{1}{6}) + (\frac{5}{6})^3(\frac{1}{6}) \\ &= 0.5177 \end{aligned}$$

Another way is to use the cumulative distribution function:

$$\begin{aligned} & P(X \leq x) = 1 - (1-p)^x \\ \implies & P(X \leq 4) = 1 - (\frac{5}{6})^4 \\ & \qquad = 0.5177. \end{aligned}$$

(c) 0.0260. This could be answered using the brute force method: $P(X > 20) = 1 - P(X \leq 20) = 1 - [P(X = 1) + P(X = 2) + \ldots + P(X = 20)]$. However, this would require calculating 20 geometric probabilities which would be very cumbersome without using software. It is easier to answer this question if we recognize that the first winner will come after the 20th cup *if the first 20 cups have no winners.* So $P(X > 20) = (\frac{5}{6})^{20} = 0.0260$.

(d) $\mu = \frac{1}{p} = \frac{1}{1/6} = 6$.

30. (a) The mean of the geometric distribution is $\frac{1}{p}$. *True.*

(b) The most likely value of a geometric random variable is always 1. *True. The most likely value is always 1, the next most likely value is always 2, and so on.*

(c) If $p > 0.5$, then $P(X = 2) < P(X = 3)$. *False. Geometric probabilities are always decreasing in x for $x = 1, 2, \ldots$, so $P(X = 2) > P(X = 3)$.*

(d) The variance of the geometric distribution is always 1. *False.* $\sigma^2 = \frac{1-p}{p^2}$.

(e) A geometric random variable can take on a countably infinite number of possible values. *True. X takes on the values 1, 2, ...*

(f) If X has a geometric distribution, then $Y = \frac{1}{X}$ also has a geometric distribution. *False. The possible values of Y would be 1, $\frac{1}{2}$, $\frac{1}{3}$, ... These are not the values that a geometric random variable takes on.*

5.9 The Negative Binomial Distribution

31. Let X represent the number of attempts required to chip-in 4 times. X has the negative binomial distribution with parameters $r = 4$ and $p = 0.05$. The probability mass function of the negative binomial distribution is: $P(X = x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}$.

(a) $P(X=20) = \binom{20-1}{4-1} 0.05^4 (1-0.05)^{16} = 0.00267$.

(b) $P(X=20)+P(X=21) = \binom{20-1}{4-1} 0.05^4 (1-0.05)^{16} + \binom{21-1}{4-1} 0.05^4 (1-0.05)^{17} = 0.00267 + 0.00298 = 0.00564$.

(c) $\mu = \frac{r}{p} = \frac{4}{0.05} = 80$.

(d) $\sigma = \sqrt{\frac{r(1-p)}{p^2}} = \sqrt{\frac{4 \cdot 0.95}{0.05^2}} = 38.99$.

(e) It will take more than 100 attempts to chip-in 4 times if the first 100 attempts have fewer than 4 chip-ins. Let Y represent the number of chip-ins in the first 100 attempts. Then we need to find $P(Y < 4)$, where Y has the binomial distribution with $n = 100$ and $p = 0.05$.

$$\begin{aligned} P(Y < 4) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\ &= 0.0059 + 0.0312 + 0.0812 + 0.1396 \\ &= 0.2578 \end{aligned}$$

This implies that $P(X > 100) = 0.2578$.

32. Let X represent the number of doors required to make 3 sales. Then X has the negative binomial distribution with parameters $r = 3$ and $p = 0.015$. The probability mass function of the negative binomial distribution is: $P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$.

(a) $P(X=50) = \binom{50-1}{3-1} 0.015^3 (1-0.015)^{47} = 0.00195$.

(b) $P(X=200) = \binom{200-1}{3-1} 0.015^3 (1-0.015)^{197} = 0.00339$.

(c) $\mu = \frac{r}{p} = \frac{3}{0.015} = 200$.

(d) $\sigma = \sqrt{\frac{r(1-p)}{p^2}} = \sqrt{\frac{3 \cdot 0.985}{0.015^2}} = 114.60$.

(e) It will take more than 500 attempts to make 3 sales if the first 500 attempts have fewer than 3 sales. Let Y represent the number of sales in the first 500 attempts. Then we need to find $P(Y < 3)$, where Y has the binomial distribution with $n = 500$ and $p = 0.015$.

$$\begin{aligned} P(Y < 3) &= P(Y = 0) + P(Y = 1) + P(Y = 2) \\ &= 0.00052 + 0.00398 + 0.01512 \\ &= 0.0196 \end{aligned}$$

This implies that $P(X > 100) = 0.0196$.

5.10 The Multinomial Distribution

33. Let X_1, X_2, and X_3 be the number of people with arches, loops, and whorls, respectively. Then X_1, X_2, and X_3 have the multinomial distribution with parameters $n = 10$ and $p_1 = 0.12$, $p_2 = 0.49$, and $p_3 = 0.39$.

(a) $P(X_1 = 2, X_2 = 3, X_3 = 5) = \frac{10!}{2!3!5!} 0.12^2 0.49^3 0.39^5 = 0.0385.$

(b) $P(X_1 = 4, X_2 = 4, X_3 = 2) = \frac{10!}{4!4!2!} 0.12^4 0.49^4 0.39^2 = 0.0057.$

(c) Since we are counting the number with arches, we will label arches a *success* and non-arches (loops or whorls) a *failure*. The number with arches (X_1) will have a binomial distribution with $n = 10$ and $p = 0.12$.
$P(X_1 = 1) = \binom{10}{1} 0.12^1 (1 - 0.12)^9 = 0.3798.$

(d) The number with arches has a binomial distribution with parameters n and p, so $E(X_1) = np = 10 \times 0.12 = 1.2$.

34. Let X_1, X_2, and X_3 be the number of times the ball lands in red, black, and green slots, respectively. Then X_1, X_2, and X_3 have the multinomial distribution with parameters $n = 20$ and $p_1 = \frac{18}{37}$, $p_2 = \frac{18}{37}$, and $p_3 = \frac{1}{37}$.

(a) $P(X_1 = 8, X_2 = 10, X_3 = 2) = \frac{20!}{8!10!2!} (\frac{18}{37})^8 (\frac{18}{37})^{10} (\frac{1}{37})^2 = 0.0141.$

(b) $P(X_1 = 14, X_2 = 3, X_3 = 3) = \frac{20!}{14!3!3!} (\frac{18}{37})^{14} (\frac{18}{37})^3 (\frac{1}{37})^3 = 0.000073.$

(c) Since we are counting the number of times the ball lands in a red slot, we will label the red slots a *success* and non-red slots (green or black) a *failure*. The number of times the ball lands in a red slot (X_1) will have a binomial distribution with $n = 20$ and $p = \frac{18}{37}$.
$P(X_1 = 11) = \binom{20}{11} (\frac{18}{37})^1 (1 - \frac{18}{37})^9 = 0.1506.$

(d) The number of times the ball lands in a red slot has binomial distribution with parameters $n = 20$ and $p = \frac{18}{37}$, so $E(X_1) = np = 20 \times \frac{18}{37} = \frac{360}{37} \approx 9.73$.

35. All of the statements are true.

5.11 Chapter Exercises

5.11.1 Basic Calculations

36. (a) $c = 0.05$, since the probabilities must sum to 1. So the distribution is:

x	-5	10	50	100
$p(x)$	0.05	0.05	0.40	0.50

(b) $P(X > 21) = P(X = 50) + P(X = 100) = 0.40 + 0.50 = 0.90.$

(c) $P(X > 21 | X < 60) = \frac{P(X > 21 \cap X < 60)}{P(X < 60)} = \frac{P(X = 50)}{P(X < 60)} = \frac{0.40}{0.05 + 0.05 + 0.40} =$ 0.80.

37. (a) 115. The missing probability must be 0.4. $E(X) = 10 \cdot 0.1 + 20 \cdot 0.2 + 30 \cdot 0.3 + x \cdot 0.4 = 60 \implies x = 115.$

38. (a) 0.1 (since the probabilities must sum to 1).
    (b) $P(X > 11) = P(X = 12) + P(X = 13) = 0.2 + 0.1 = 0.3$.
    (c) $E(X) = \sum xp(x) = 10 \cdot 0.4 + 11 \cdot 0.3 + 12 \cdot 0.2 + 13 \cdot 0.1 = 11$.
    (d) $\sigma^2 = 1$. $E(X^2) = \sum x^2 p(x) = 10^2 \cdot 0.4 + 11^2 \cdot 0.3 + 12^2 \cdot 0.2 + 13^2 \cdot 0.1 = 122$. $\sigma^2 = E(X^2) - [E(X)]^2 = 122 - 11^2 = 1$.

39. (a) 0.2304. Let X represent the number of 10s selected. Then X has a binomial distribution with $n = 5$ and $p = 0.4$. $P(X = 3) = \binom{5}{3} 0.4^3 0.6^2 = 0.2304$.
    (b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.6^5 = 0.92224$.
    (c) This will occur if there are no 11s in the first six values, which has a probability of $(1 - 0.3)^6 = 0.1176$.

40. Let X be the number of 10s selected. Then X has a binomial distribution with $n = 20$ and $p = 0.4$.
    (a) $P(X < 8) = 0.416$. (This requires summing 8 binomial probabilities, so it is best carried out using software.)
    (b) $P(2 \leq X \leq 6) = 0.249$.
    (c) $E(X) = np = 20 \cdot 0.40 = 8$.

41. Both X and Y are binomial random variables. Binomial random variables have a mean of np and a variance of $np(1 - p)$.
    (a) $E(X + Y) = E(X) + E(Y) = 20 \times 0.50 + 18 \times \frac{1}{6} = 13$.
    (b) $\sigma_X^2 = 20 \times 0.5 \times (1 - 0.5) = 5$, $\sigma_Y^2 = 18 \times \frac{1}{6} \times \frac{5}{6} = 2.5$. $\sigma_{X-Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{5 + 2.5} \approx 2.739$.

42. (a) $P(X > 1) = P(X = 10) + P(X = 100) = 0.01 + 0.001 = 0.011$.
    (b) The only way to get a total profit of \$100 is to have one \$100 ticket and one \$0 ticket. The probability of this is $0.689 \cdot 0.001 + 0.001 \cdot 0.689 = 0.001378$.
    (c) $E(X) = \sum xp(x) = 0 \cdot 0.689 + 1 \cdot 0.30 + 10 \cdot 0.010 + 100 \cdot 0.001 = 0.5$.
    (d) $\sigma^2 = \sum (x - \mu)^2 p(x) = (0 - 0.5)^2 \cdot 0.689 + (1 - 0.5)^2 \cdot 0.30 + (10 - 0.5)^2 \cdot 0.010 + (100 - 0.5)^2 \cdot 0.001 = 11.05$. $\sigma = \sqrt{11.05} = 3.324$.
    (e) $P(X = 100 | X > 0) = \frac{P(X=100)}{P(X=1)+P(X=10)+P(X=100)} = \frac{0.001}{0.3+0.01+0.001} = 0.003215$.
    (f) $1000 \cdot 0.5 = \$500$.

5.11.2 Concepts

43.

$$\begin{aligned}
E[(X-\mu)^2] &= E(X^2 - 2X\mu + \mu^2) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - 2\mu E(X) - \mu^2 \quad \text{(Since } \mu \text{ is a constant)} \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}$$

44. No, that statement is false. A variable is not independent of itself.

45. All of the statements are false.

46. (a) $E(X) = 15 \cdot 0.2 = 3$, $E(Y) = 6$, and $E(Z) = 5$.
$E(X + Y + Z) = E(X) + E(Y) + E(Z) = 3 + 6 + 5 = 14$.
(b) $\sigma_X^2 = np(1-p) = 15 \cdot 0.2 \cdot 0.8 = 2.4$, $\sigma_Y^2 = 6$, $\sigma_Z^2 = 18^2 = 324$.
$\sigma_{X+Y-Z} = \sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2} = \sqrt{2.4 + 6 + 324} = \sqrt{332.4} = 18.232$.
(c) Yes. X is a binomial random variable and cannot take on negative values, and Y is a Poisson random variable and cannot take on negative values. But Z has an unknown distribution, so it could possibly take on negative values. Since $X + Y$ could be 0, $X + Y - Z$ could possibly take on negative values.

47. (e) (None of the above). The number of children in the selected area will be 0 or 1 (since the children spread themselves out). The number of children in the area would have a Bernoulli distribution, but the precise value of p is a little tricky to determine.

48. (c). If the trials were independent, then X would have a binomial distribution. However, in this scenario the trials are not independent, since if Tom makes a mistake then he is more likely to make more mistakes. X is a count variable with a minimum of 0 and a maximum of 20, and it is thus a discrete random variable, but its exact distribution is impossible to determine without further information.

49. (a) A discrete random variable can take on a countable number of possible values. *True.*
(b) A discrete random variable always represents a count. *False.*
(c) The expected value of a discrete random variable must equal one of the variable's possible values. *False.*

(d) The binomial distribution has two parameters: n and p. *True.*

(e) A binomial random variable can take on one of $n+1$ possible values. *True. (A binomial random variable takes on the values* $0, 1, 2, \ldots, n$.*)*

50. (a) The variance and standard deviation of a discrete random variable cannot be equal. *False. (They may both equal 1.)*

(b) The mean of a random variable cannot be negative. *False.*

(c) The standard deviation of a random variable cannot be negative. *True.*

(d) If X is a discrete random variable, then $Y = 2X + 3$ is a discrete random variable. *True.*

51. (a) In a binomial setting, the probability of success usually changes from trial to trial. *False. (In a binomial setting, the probability of success stays constant from trial to trial.)*

(b) The mean of a Poisson random variable can be negative. *False.* ($\mu = \lambda$ *must be positive.)*

(c) If X represents the number of students in a randomly selected 100 m^2 area of a university campus at 3:00 pm on a Tuesday afternoon, then X has a Poisson distribution. *False. (Students are not distributed randomly—they tend to group together.)*

(d) If $p > 0$, then the variance of a binomial random variable is less than its mean. *True.*

(e) If X is a binomial random variable with $n = 10$ and $p = 0.1$, and Y is a Poisson random variable with $\lambda = 2$, then $X - Y$ has a binomial distribution. *False.* ($X - Y$ *can take on negative values so it cannot possibly have a binomial distribution.)*

5.11.3 Applications

52. (a) $E(X) = 1 \cdot 0.38 + 2 \cdot 0.52 + 3 \cdot 0.05 + 4 \cdot 0.05 = 1.77$.

(b) $\sigma^2 = (1-1.77)^2 \cdot 0.38 + (2-1.77)^2 \cdot 0.52 + (3-1.77)^2 \cdot 0.05 + (4-1.77)^2 \cdot 0.05 = 0.5771$. $\sigma = \sqrt{0.5771} = 0.7597$.

(c) $E(ln(X)) = ln(1) \cdot 0.38 + ln(2) \cdot 0.52 + ln(3) \cdot 0.05 + ln(4) \cdot 0.05 = 0.48468$.

(d) $E(X^2) = 1^2 \cdot 0.38 + 2^2 \cdot 0.52 + 3^2 \cdot 0.05 + 4^2 \cdot 0.05 = 3.71$.

53. (a) Let Y represent the number of randomly selected numbers that match the ticket. Then Y has a hypergeometric distribution with $a = 6$, $N = 70$, and $n = 20$.

$P(Y = 5) = \frac{\binom{6}{5}\binom{64}{15}}{\binom{70}{20}} = 0.0059123$. $P(Y = 6) = \frac{\binom{6}{6}\binom{64}{14}}{\binom{70}{20}} = 0.0002956$.

$P(Y \leq 4) = 1 - [P(Y = 5) + P(Y = 6)] = 1 - (0.0059123 + 0.0002956) = 0.9937921$.

(b) Let X represent the payout on the ticket.
$E(X) = \sum xp(x) = 0 \cdot 0.9937921 + 25 \cdot 0.0059123 + 1000 \cdot 0.0002956 = 0.4434$.

(c) $E(X^2) = \sum x^2 p(x) = 0^2 \cdot 0.9937921 + 25^2 \cdot 0.0059123 + 1000^2 \cdot 0.0002956 = 299.2952$. $\sigma^2 = E(X^2) - [E(X)]^2 = 299.2952 - 0.4434^2 = 299.0986$. $\sigma = \sqrt{299.0986} = 17.29$.

(d) A player is paying \$1 for the ticket and getting back \$0.4434 on average, so the government is keeping about 56 cents on average.

54. (a) Let X represent the number of tickets that win \$25. Then X has a binomial distribution with parameters $n = 50$ and $p = 0.0059123$.
$P(X = 1) = \binom{50}{1} 0.0059123^1 (1 - 0.0059123)^{49} = 0.2211$.

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{50}{0} 0.0059123^0 (1 - 0.0059123)^{50} = 1 - 0.7434 = 0.2566$.

(c) Let W represent the number of tickets that win a prize (either \$25 or \$1000). Then W has a binomial distribution with parameters $n = 50$ and $p = 0.0059123 + 0.0002956 = 0.0062079$. $P(W \geq 1) = 1 - P(W = 0) = 1 - \binom{50}{0} 0.0062079^0 \cdot 0.9937921^{50} = 1 - 0.9937921^{50} = 1 - 0.7324 = 0.2676$.

(d) $E(W) = np = 50 \cdot 0.0062079 = 0.3104$.

(e) $\sigma_W = \sqrt{np(1-p)} = \sqrt{50 \cdot 0.0062079(1 - 0.0062079)} = 0.5554$.

55. (a) $5000 \times 0.98 + 100000 \times 0.02 = 6900$.

(b) Taking the gamble saves you \$100 on average.

(c) We are often willing to accept a reduced expectation in order to reduce the variability. (For example, when we purchase insurance.) Here, perhaps losing the wager would put you in a bad financial position, which is likely something you want to avoid. (And there is always the possibility that the salesman is running some sort of con game on you. It is, of course, probably best to avoid propositions like this.)

56. (a) $1000 - 25000 \cdot 0.01 = \$750$.

(b) Let X represent the insurance company's payout. The distribution of X is:

x	0	25000
$p(x)$	0.99	0.01

$E(X) = 0 \cdot 0.99 + 25000 \cdot 0.01 = 250$
$E(X^2) = 0^2 \cdot 0.99 + 25000^2 \cdot 0.01 = 6250000$
$\sigma^2 = E(X^2) - [E(X)]^2) = 6250000 - 250^2 = 6187500$.
$\sigma = \sqrt{6187500} = 2487.469$.

(c) $1000 - 25000 \cdot p = 0 \implies p = \frac{1000}{25000} = 0.04$.

57. (a) The number of siblings that are an HLA-identical match has a binomial distribution with $n = 6$ and $p = \frac{1}{4}$.

(b) Let X represent the number of siblings that are an HLA-identical match. $P(X \geq 1) = 1 - P(X = 0) = 1 - (\frac{3}{4})^6 = 0.822$.

(c) $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - (0.1780 + 0.3560) = 0.466$.

(d) $\mu = np = 6 \times \frac{1}{4} = 1.5$.

58. Let X represent the number of completed surveys. Then X has a binomial distribution with parameters $n = 100$ and $p = 0.09$.

(a) $P(X = 8) = \binom{100}{8} 0.09^8 0.91^{92} = 0.1366$.

(b) 0.399. $P(X = 8)+P(X = 9)+P(X = 10) = \binom{100}{8} 0.09^8 0.91^{92}+\binom{100}{9} 0.09^9 0.91^{91}+\binom{100}{10} 0.09^{10} 0.91^{90} = 0.1366 + 0.1381 + 0.1243 = 0.399$.

(c) $P(X \geq 2) = 1-[P(X = 0)+P(X = 1)] = 1-(0.00008+0.00079) = 0.99913$.

(d) $\mu = np = 100 \cdot 0.09 = 9$.

(e) $\sigma = \sqrt{np(1-p)} = \sqrt{100 \cdot 0.09 \cdot 0.91} = 2.862$.

59. Let X represent the number of calls needed to get 4 completed surveys. Then X has the negative binomial distribution with parameters $r = 4$ and $p = 0.09$. The probability mass function of the negative binomial distribution is: $P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$.

(a) $P(X = 50) = \binom{50-1}{4-1} 0.09^4 (1-0.09)^{46} = 0.0158$.

(b) $P(X = 50)+P(X = 51) = \binom{50-1}{4-1} 0.09^4 (1-0.09)^{46} + \binom{51-1}{4-1} 0.09^4 (1-0.09)^{47} = 0.0158 + 0.0153 = 0.0311$.

(c) $\mu = \frac{r}{p} = \frac{4}{0.09} = 44.4$.

(d) $\sigma = \sqrt{\frac{r(1-p)}{p^2}} = \sqrt{\frac{4 \cdot 0.91}{0.09^2}} = 21.20$.

(e) It will take more than 100 calls to get 4 completed surveys if the first 100 calls result in fewer than 4 completed surveys. Let Y represent the number of completed surveys in the first 100 calls. Then we need to find $P(Y < 4)$, where Y has the binomial distribution with $n = 100$ and $p = 0.09$.

$$\begin{aligned} P(Y < 4) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\ &= 0.00008 + 0.00079 + 0.00388 + 0.01254 \\ &= 0.0173 \end{aligned}$$

This implies that $P(X > 100) = 0.0173$.

60. Let X represent the number of chickens that have detectable levels of salmonella. Then, assuming we have a random and independent sample of chickens, X has a binomial distribution with parameters $n = 12$ and $p = 0.14$.

(a) $P(X = 2) = \binom{12}{2} 0.14^2 0.86^{10} = 0.286$.

(b) 0.7697.

$$\begin{aligned}
P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
&= \binom{12}{0} 0.14^0 0.86^{12} + \binom{12}{1} 0.14^1 0.86^{11} + \binom{12}{2} 0.14^2 0.86^{10} \\
&= 0.1637 + 0.3197 + 0.2863 \\
&= 0.7697
\end{aligned}$$

(c) Let Y represent the number of chickens with detectable levels of salmonella. Assuming this retailer's claim is true, Y has a binomial distribution with $n = 20$ and $p = 0.02$. $P(Y \geq 3) = 1 - P(X \leq 2) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)] = 0.007$. This yields strong evidence against the company's claim.

61. Let X represent the number of people in the sample that have given blood at some point.

(a) X has a binomial distribution with $n = 5$ and $p = 0.10$.
(b) $P(X = 1) = 0.328$.
(c) $P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 0.0815$.
(d) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.9914$.
(e) $P(X = 1|X \geq 1) = \frac{P(X=1 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X=1)}{P(X \geq 1)} = \frac{0.32805}{0.40951} = 0.8011$.

62. (a) The number of tickets that win a cash prize follows a binomial distribution with $n = 156$ and $p = 0.03$.
(b) $E(X) = np = 156 \cdot 0.03 = 4.68$.
(c) $P(X = 0) = 0.97^{156} = 0.0086$.
(d) This is a complicated question, without an easy answer, and there are many factors to take into account (for example, how trustworthy you believe your friend is). But if your friend was indeed purchasing the tickets and (theoretically) sharing the cash prizes, there is a less than 1% chance of not winning a prize. You may not want to end a friendship over this, but it is a little fishy.

63. Let X represent the number of crashes in the next year. Since the fatal crashes can be thought of as occurring randomly and independently with a mean of 1.1, X has (approximately) a Poisson distribution with $\lambda = 1.1$.

(a) $P(X = 0) = \frac{1.1^0 e^{-1.1}}{0!} = 0.3329$.
(b) $P(X = 2) + P(X = 3) = \frac{1.1^2 e^{-1.1}}{2!} + \frac{1.1^3 e^{-1.1}}{3!} = 0.2014 + 0.0738 = 0.2752$.
(c) $P(X > 4) = 1 - P(X \leq 4) = 0.0054$.
(d) $\sigma = \sqrt{\lambda} = \sqrt{1.1} = 1.049$.

64. Let X represent the number of crashes in the next 2 year period. Since the fatal

crashes can be thought of as occurring randomly and independently with a mean of 1.1 per year, X has (approximately) a Poisson distribution with $\lambda = 2 \times 1.1 = 2.2$.

(a) $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{2.2^0 e^{-2.2}}{0!} = 1 - 0.1108 = 0.8892$.
(b) $P(X > 1) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)] = 1 - (0.1108 + 0.2438) = 0.6454$.
(c) $\mu = \lambda = 2.2$.
(d) $\sigma = \sqrt{\lambda} = \sqrt{2.2} = 1.48$.

65. (a) If the mean number of impacts per million years is 2, then the mean number of impacts in a 100,000 year period is $\mu = \lambda = \frac{2}{10} = 0.20$.
$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{0.2^0 e^{-0.2}}{0!} = 1 - 0.8187 = 0.1813$.
(b) If the mean number of impacts per million years is 2, then the mean number of impacts in a 10,000 year period is $\mu = \lambda = \frac{2}{100} = 0.02$.
$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{0.02^0 e^{-0.02}}{0!} = 1 - 0.9802 = 0.0198$.

66. (a) The number of round/yellow plants has a binomial distribution with $n = 50$ and $p = \frac{9}{16}$.
(b) $\mu = np = 50 \cdot \frac{9}{16} = 28.125$.
(c) $\sigma = \sqrt{np(1-p)} = \sqrt{50(\frac{9}{16})(1 - \frac{9}{16})} = 3.5078$.
(d) The number of wrinkled/green plants has a binomial distribution with $n = 50$ and $p = \frac{1}{16}$. $P(X = 4) = \binom{50}{4}(\frac{1}{16})^4(\frac{15}{16})^{46} = 0.1805$.
(e) $\frac{50!}{28!12!6!4!}(\frac{9}{16})^{28}(\frac{3}{16})^{12}(\frac{3}{16})^6(\frac{1}{16})^4 = 0.00152$. (The distribution of the number of plants with the 4 phenotypes is the multinomial distribution with $n = 50$, $p_1 = \frac{9}{16}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{3}{16}$, and $p_4 = \frac{1}{16}$.)

5.11.4 Extra Practice Questions

67. (a) $\mu = np \implies 2.4 = 10 \cdot p \implies p = 0.24$.
(b) $P(X = 2) = 0.2885$.
(c) $\sigma = \sqrt{np(1-p)} = \sqrt{10 \cdot 0.24 \cdot (1 - 0.24)} = \sqrt{1.824} = 1.350$.

68. X is a binomial random variable with $n = 2$, $p = \frac{2}{3}$.

(a) $\mu = np = 2 \cdot \frac{2}{3} = \frac{4}{3}$.
(b) $\sigma = \sqrt{np(1-p)} = \sqrt{2 \cdot \frac{2}{3} \cdot \frac{1}{3}} = \frac{2}{3}$.

69. (a) The number of people in the sample that are over 60 years old will follow a binomial distribution with $n = 20$ and $p = 0.178$. Let X represent this number.
(b) $P(X \geq 2) = 0.89426$.

(c) $P(X = 2) = 0.1767$.
(d) $P(X = 0) = (1 - 0.178)^{20} = 0.01984$.
(e) $E(X) = np = 20 \cdot 0.178 = 3.56$.

70. (a) The number of people in the sample that voted in the last federal election has a binomial distribution with $n = 5$ and $p = 0.60$. Let X represent this number.
(b) $P(X = 4) = 0.2592$.
(c) $P(X \geq 4) = P(X = 4) + P(X = 5) = 0.3370$.
(d) $P(X = 5|X \geq 4) = \frac{P(X=5 \cap X \geq 4)}{P(X \geq 4)} = \frac{P(X=5)}{P(X \geq 4)} = \frac{0.07776}{0.33696} = 0.2308$.
(e) $E(X) = np = 5 \cdot 0.60 = 3.0$.

71. Let X be the number of people that are out of work. Then X has a binomial distribution with $n = 14$ and $p = 0.076$.

(a) $P(X = 3) = 0.0670$.
(b) $P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] = 0.08496$.
(c) 1. (If there are at least 2 out of work, there must be at least 1 out of work.)
(d) 0. (If there are at least 3 out of work, then there is more than 1 out of work.)

72. The number of cars that fail the test will have a binomial distribution with $n = 10$ and $p = 0.35$.

(a) $P(X = 2) = 0.17565$.
(b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.2616$.
(c) $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.2616 = 0.7384$.
(d) $E(X) = np = 10 \cdot 0.35 = 3.5$.
(e) $\sigma = \sqrt{np(1-p)} = \sqrt{10 \cdot 0.35 \cdot 0.65} = 1.508$.
(f) Not in and of itself, but something is going on. If this was a random selection of 10 cars that are at least 13 years old, then the probability of failing all of them should be approximately $0.35^{10} = 0.000028$. This is quite unlikely, so there is a concern that the garage might be inappropriately failing cars for some reason. But there are other possible explanations. For example, perhaps this garage has an excellent reputation for fixing cars that are almost ready for the junk heap, so they might tend to test cars that are in worse shape.

73. (a) The number of decays in a one-second period is assumed to follow a Poisson distribution with $\lambda = 0.8$, so: $P(X = 2) = \dfrac{0.8^2 e^{-0.8}}{2!} = 0.1438$.
(b) The number of events in a three-second period is assumed to follow a Poisson distribution with $\lambda = 3 \times 0.8 = 2.4$, so: $P(X = 2) = \dfrac{2.4^2 e^{-2.4}}{2!} = 0.2613$.

(c) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{2.4^0 e^{-2.4}}{0!} + \frac{2.4^1 e^{-2.4}}{1!} + \frac{2.4^2 e^{-2.4}}{2!} = 0.5697$.

74. Since the events are occurring randomly and independently of one another, it is reasonable to assume that the number of events in an 8 hour period follows a Poisson distribution with $\lambda = 8 \times 0.2 = 1.6$.

(a) Let X represent the number of log jams in an 8 hour period. $P(X = 3) = 0.1378$.

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.2019 = 0.7981$.

(c) 0.1582. Let X represent the number of log jams in an hour. The probability there is at least one log jam is $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.8187 = 0.1813$ (calculated using the Poisson pmf with $\lambda = 0.2$).
Let Y represent the number of hours in the 12 randomly selected hours that have at least one log jam. Then Y has a binomial distribution with $n = 12$ and $p = 0.1813$. $P(Y \geq 4) = 1 - P(Y \leq 3) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3)] = 1 - 0.8418 = 0.1582$.

75. (a) 0.999955. Let X represent the number of planes that arrive in the two hour period. $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.000045 = 0.999955$ (calculated using the Poisson pmf with $\lambda = 2 \times 5 = 10$).

(b) 0.00017. Let X represent the number of planes that arrive in a 3 hour period. $P(X = 3) = 0.00017$ (calculated using the Poisson pmf with $\lambda = 3 \times 5 = 15$).

(c) 0.0821. Let X represent the number of planes that arrive in a half-hour period. $P(X = 0) = 0.0821$ (calculated using the Poisson pmf with $\lambda = 0.5 \times 5 = 2.5$).

(d) $\sqrt{8 \times 5} = 6.325$ (since for a Poisson random variable, $\sigma = \sqrt{\lambda}$).

76. (a) Under the given assumptions, the number that crash has a binomial distribution with $n = 10$ and $p = 0.01$. For the remaining questions, let X represent the number of planes that crash.

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.90438 = 0.0956$ (calculated using the binomial pmf with $n = 10$ and $p = 0.01$).

(c) $P(X = 1) = 0.09135$ (calculated using the binomial pmf with $n = 10$ and $p = 0.01$).

(d) $P(X = 10) = 10^{-20} = 0.00000000000000000001$.

(e) $\sigma = \sqrt{np(1-p)} = \sqrt{10 \cdot 0.01 \cdot (1 - 0.01)} = 0.3146$.

77. (a) The number that need to be selected in order to get the first doctor has a geometric distribution with $p = 0.002$. Let X represent this number. $P(X = 4) = (1 - 0.002)^3 \times 0.002 = 0.001988$.

(b) If the first 1000 people are not doctors, then the first doctor will occur after the first 1000 people sampled. $P(X > 1000) = (1 - 0.002)^{1000} = 0.1351$.

(c) The number of doctors in a random sample of 80 Canadians has (approximately) a binomial distribution with $n = 80$ and $p = 0.002$. Let Y represent this number. $P(Y = 1) = 0.13659$.

(d) $P(Y > 2) = 1 - P(Y \leq 2) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)] = 1 - 0.9994 = 0.0005858$ (calculated using the binomial pmf with $n = 80$ and $p = 0.002$).

78. (a) The number of welds needed to get the first unsuccessful weld has a geometric distribution with $p = 0.10$. For the following questions, let X represent this random variable.

(b) $P(X = 3) = 0.9^2 \cdot 0.1 = 0.081$.

(c)

$$\begin{aligned} P(X \leq 3) &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.9^0 \cdot 0.1 + 0.9^1 \cdot 0.1 + 0.9^2 \cdot 0.1 \\ &= 0.271 \end{aligned}$$

(d) The first unsuccessful weld will occur after the 13th trial if the first 13 trials are successful. So $P(X > 13) = 0.9^{13} = 0.254$.

79. (a) The number of welds needed to get the third unsuccessful weld has a negative binomial distribution with $r = 3$ and $p = 0.10$. Let X represent this random variable.
$P(X = 22) = \binom{21}{2} 0.1^3 0.9^{19} = 0.02837$

(b) Let Y represent the number of unsuccessful welds. Then Y has a binomial distribution with $n = 20$ and $p = 0.10$. $P(X = 3) = 0.1901$.

(c) $P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)] = 1 - 0.6769 = 0.3231$.

(d) $\mu = np = 20 \times 0.1 = 2$.

80. (a) Let X represent the number of shipments needed to get the first rejected shipment. Then X has a geometric distribution with $p = 0.1$. $P(X = 10) = 0.96^9 \cdot 0.04 = 0.0277$.

(b) The first rejected shipment will occur after the 10th shipment if the first 10 are not rejected. $P(X > 10) = 0.96^{10} = 0.6648$.

(c) Let Y represent the number of rejected shipments. Then Y has a binomial distribution with $n = 60$ and $p = 0.04$. $P(Y \leq 1) = P(Y = 0) + P(Y = 1) = 0.0864 + 0.2159 = 0.3022$.

81. (a) The number of men in the sample follows a hypergeometric distribution. To get exactly 4 men, we must choose 4 men from 55, and 1 woman from 30: $P(X = 4) = \frac{\binom{55}{4}\binom{30}{1}}{\binom{85}{5}} = 0.3119$.

(b) To approximate the hypergeometric distribution in this example, we can use the binomial distribution with $n = 5$ and $p = \frac{55}{55+30} = \frac{11}{17}$. Using the binomial pmf: $P(X = 4) \approx 0.3093$. Note that this is not too far from the correct probability of 0.3119 (found in 81a), since we are sampling only a small fraction of the population.

82. Let X represent the number of dandelions in the randomly selected area. Then X has a Poisson distribution with $\lambda = 4$.

(a) $P(X = 2) = 0.1465$.
(b) $P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] = 1 - 0.2381 = 0.7619$.
(c) $P(X = 3|X \geq 1) = \frac{P(X=3 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X=3)}{P(X \geq 1)} = \frac{0.1953668}{0.9816844} = 0.1990$.
(d) $\sigma = \sqrt{\lambda} = \sqrt{4} = 2$.

83. (a) Let X represent the number of implants needed to get the first unsuccessful one. Then X has a geometric distribution with $p = 0.05$. $P(X = 4) = 0.95^3 \cdot 0.05 = 0.0429$.
(b) In order for the first unsuccessful implant to occur after the 11th trial, we need the first 11 trials to be successful. $P(X > 11) = 0.95^{11} = 0.5688$.
(c) Let Y represent the number of trials needed to get the third unsuccessful implant. Then Y has the negative binomial distribution with $r = 3$ and $p = 0.05$. $P(Y = 28) = \binom{27}{2} 0.05^3 0.95^{25} = 0.0122$.

84. (a) Here we need to pick 2 footballs and 4 balls that are either soccer balls or volleyballs:
$P(2 \text{ footballs}) = \frac{\binom{15}{2}\binom{28}{4}}{\binom{43}{6}} = 0.3526$.
(b) $\mu = n \cdot \frac{a}{N} = 6 \cdot \frac{15}{43} = 2.093$.
(c) $P(0 \text{ soccer balls}) + P(1 \text{ soccer ball}) = \frac{\binom{20}{0}\binom{23}{6}}{\binom{43}{6}} + \frac{\binom{20}{1}\binom{23}{5}}{\binom{43}{6}} = 0.01656 + 0.11039 = 0.1269$.
(d) $\frac{\binom{15}{3}\binom{20}{2}\binom{8}{1}}{\binom{43}{6}} = 0.1134$.

85. (a) Since the sampling is done *without* replacement, the number of computer engineers in the group of 6 will follow a hypergeometric distribution. We must select 3 computer engineers (from 7 total) and 3 non-computer engineers (from $5 + 5 + 3 = 13$ total). $P(X = 3) = \frac{\binom{7}{3}\binom{13}{3}}{\binom{20}{6}} = 0.2583$.
(b) Here we must select 2 biomedical engineers (from 5), 2 computer engineers (from 7), 1 mechanical engineer (from 5) and 1 water resource engineer (from

3). The probability of this is: $\dfrac{\binom{5}{2}\binom{7}{2}\binom{5}{1}\binom{3}{1}}{\binom{20}{6}} = 0.0812$.

86. Here we need to pick 1 Honda vehicle (from 4), 1 Ford (from 6), 1 Toyota (from 3), and 2 Chrysler (from 7). We can calculate the probability of this event using the *multivariate hypergeometric* distribution: $\dfrac{\binom{4}{1}\binom{6}{1}\binom{3}{1}\binom{7}{2}}{\binom{20}{5}} = 0.09752$.

87. $\frac{5!}{3!1!1!0!}0.4^3 0.3^1 0.2^1 0.1^0 = 0.0768$.

Chapter 6

Continuous Random Variables and Continuous Probability Distributions

J.B.'s strongly suggested exercises: 1, 5, 6, 8, 9, 10, 11, 13, 14, 15, 21, 23, 24, 27

6.1 Introduction

6.2 Properties of Continuous Probability Distributions

1. (a) For continuous random variables, probabilities correspond to areas under the density curve. *True.*
   (b) The area under any continuous probability distribution is 1. *True.*
   (c) A probability density function can take on negative values. *False.* $f(x) \geq 0$ *everywhere.*
   (d) A probability density function can take on values greater than 1. *True. The area under the entire curve is 1, but the height of the curve can be greater than 1.*
   (e) $f(x) = P(X = x)$ for all x. *False. The value of the pdf (the height of the curve at point x) is* not *a probability.*
   (f) $f(x) = \mu$ whenever the distribution is symmetric. *False.*

6.2.1 An Example Involving Integration

2. (a) $c = \frac{1}{9}$. For any continuous probability distribution, $\int_{-\infty}^{\infty} f(x)dx = 1$. For this distribution, $\int_0^3 cx^2 dx = 1 \implies \frac{cx^3}{3}|_0^3 = 1 \implies 9c = 1 \implies c = \frac{1}{9}$.
   (b) $P(X > 2) = \int_2^3 \frac{x^2}{9} dx = \frac{x^3}{27}|_2^3 = \frac{3^3}{27} - \frac{2^3}{27} = \frac{19}{27}$.
   (c) $P(X < 1) = \int_0^1 \frac{x^2}{9} dx = \frac{x^3}{27}|_0^1 = \frac{1^3}{27} - \frac{0^3}{27} = \frac{1}{27}$.
   (d) 2.052. We need to find the value a such that $\int_0^a \frac{x^2}{9} dx = 0.32$. This implies $\frac{a^3}{27} - \frac{0^3}{27} = 0.32$. $a = \sqrt[3]{0.32 \cdot 27} = 2.052$.
   (e) 2.381. We need to find the value a such that $\int_0^a \frac{x^2}{9} dx = 0.50$. This implies $\frac{a^3}{27} - \frac{0^3}{27} = 0.50$. $a = \sqrt[3]{0.50 \cdot 27} = 2.381$.
   (f) For any continuous probability distribution, $E(X) = \int_{-\infty}^{\infty} xf(x)dx$. For this distribution, $E(X) = \int_0^3 x \cdot \frac{x^2}{9} dx = \frac{x^4}{36}|0^3 = \frac{3^4}{36} = 2.25$.

3. (a) $c = \frac{1}{9}$. For any continuous probability distribution, $\int_{-\infty}^{\infty} f(x)dx = 1$. For this distribution, $\int_4^{10} \frac{c}{x^2} dx = 1 \implies -\frac{c}{x}|_4^{10} = 1 \implies -c(\frac{1}{10} - \frac{1}{4}) = 1 \implies c = \frac{20}{3}$.
   (b) $P(8 < X < 9) = \int_8^9 \frac{20}{3} \cdot \frac{1}{x^2} dx = -\frac{20}{3} \cdot \frac{1}{x}|_8^9 = -\frac{20}{3}(\frac{1}{9} - \frac{1}{8}) = \frac{20}{216} \approx 0.093$.
   (c) We need to find the value a such that $\int_4^a \frac{20}{3} \cdot \frac{1}{x^2} dx = 0.44$. This implies $-\frac{20}{3}(\frac{1}{a} - \frac{1}{4}) = 0.44 \implies a = 5.4348$.

4. (a) $P(X > 0.22) = \int_{0.22}^{\infty} 10e^{-10x} dx = -e^{-10x}|_{0.22}^{\infty} = 0 - (-e^{-10\times 0.22}) = e^{-2.2} \approx 0.111$.
   (b) $P(X > 0.02) = \int_{0.02}^{\infty} 10e^{-10x} dx = -e^{-10x}|_{0.02}^{\infty} = 0 - (-e^{-10\times 0.02}) = e^{-0.2} \approx 0.819$.
   (c) $P(0.08 < X < 0.12) = \int_{0.08}^{0.12} 10e^{-10x} dx = -e^{-10x}|_{0.08}^{0.12} = (-e^{-10\times 0.12}) - (-e^{-10\times 0.08}) = -e^{-1.2} - (-e^{-0.8}) \approx 0.449 - 0.301 = 0.148$.
   (d) We need to find the value x such that $P(X < x) = \int_0^x 10e^{-10u} du = 0.25$. This implies $-e^{-10u}|_0^x = 0.25 \implies 1 - e^{-10x} = 0.25$. Solving for x, $x = \frac{ln(1-0.25)}{-10} = 0.02877$.
   (e) The median is the value x such that $P(X < x) = 0.5$. So we need to find the value x such that $P(X < x) = \int_0^x 10e^{-10u} du = 0.50$. This implies $-e^{-10u}|_0^x = 0.5 \implies 1 - e^{-10x} = 0.5$. Solving for x, $x = \frac{ln(1-0.5)}{-10} = 0.0693$.

6.3 The Continuous Uniform Distribution

5. (a) Since the base is $(25 - 5 = 20)$, the height of the rectangle must be $\frac{1}{25-5} = \frac{1}{20}$. So,

$$f(x) = \begin{cases} \frac{1}{20} & \text{for } 5 \leq x \leq 25 \\ 0 & \text{elsewhere} \end{cases}$$

(b) 0 (for any continuous random variable X, $P(X = a) = 0$ for all a).
(c) $(25 - 8) \times \frac{1}{20} = 0.85$. (The area under the curve to the right of 8 is 0.85.)
(d) $P(X > 8.0|X < 7.3) = \frac{P(X>8.0 \cap X<7.3)}{P(X<7.3)} = \frac{0}{P(X<7.3)} = 0$ (since an observation cannot be both greater than 8 and less than 7.3.
(e) $P(X > 8.0|X > 7.3) = \frac{P(X>8.0 \cap X>7.3)}{P(X>7.3)} = \frac{P(8.0<X<25)}{P(X>7.3)} = \frac{(25-8.0)\times\frac{1}{20}}{(25-7.3)\times\frac{1}{20}} = \frac{0.85}{0.885} = 0.96045$.
(f) 15 (since half the area lies to the left of 15, half to the right).
(g) 15. (For symmetric distributions, the mean equals the median. We could also find the mean by integration.)
(h) Let a represent the 47th percentile. The area to the left of a is 0.47 (by the definition of a percentile). So $(a - 5) \times \frac{1}{20} = 0.47 \implies a = 14.4$.
(i) $IQR = Q_3 - Q_1$. $(Q_1 - 5) \times \frac{1}{20} = 0.25 \implies Q_1 = 10$. $(Q_3 - 5) \times \frac{1}{20} = 0.75 \implies Q_3 = 20$. $IQR = 20 - 10 = 10$.

6. (a) The uniform distribution is symmetric. *True.*
(b) For any uniform distribution, the mean and median are equal. *True. The mean and median are equal in symmetric distributions.*
(c) For any uniform distribution, $Q_1 = -Q_3$. *False.* $Q_1 = -Q_3$ *only if the uniform distribution has a mean of 0.*
(d) For any uniform distribution, $Q_3 - Q_2 = Q_2 - Q_1$. *True. For symmetric distributions, the distance between the median and third quartile will be the same as the distance between the median and first quartile.*
(e) If a random variable has a uniform distribution, then it cannot take on negative values. *False.*
(f) If a random variable has a uniform distribution, then its standard deviation is greater than its mean. *False. Depending on the value of the parameters, the standard deviation can be greater than, less than, or equal to the mean.*

6.4 The Normal Distribution

7. μ and σ^2 are the parameters of the distribution (or μ and σ). μ can be any value ($-\infty < \mu < \infty$), but σ^2 must be positive ($\sigma^2 > 0$).

8. (a) The normal distribution is symmetric about μ. *True.*
(b) If X is a normally distributed random variable, then $P(X < 0) > 0$. *True. A normally distributed random variable can take on any finite value.*
(c) If X is a normally distributed random variable, then $P(X = 0) = 0$. *True. The probability a continuous random variable takes on any single value is always 0.*

(d) The standard deviation of a normal distribution cannot be greater than the mean. *False.*

(e) The mean and median of a normal distribution are always equal. *True. The normal distribution is symmetric, and the mean and median are equal in symmetric distributions.*

6.4.1 Finding Areas Under the Standard Normal Curve

9. (a) 0.6443.
   (b) $1 - 0.6443 = 0.3557$.
   (c) 0.1443.
   (d) 0.1003.
   (e) $1 - 0.1003 = 0.8997$.
   (f) 0.
   (g) 0.3997.
   (h) $P(Z > 2.41|Z > 0) = \frac{P(Z>2.41 \cap Z>0)}{P(Z>0)} = \frac{P(Z>2.41)}{P(Z>0)} = \frac{0.008}{0.5} = 0.016$.
   (i) $P(-2.41 < Z < 2.41|Z > 0) = \frac{P(-2.41<Z<2.41 \cap Z>0)}{P(Z>0)} = \frac{P(0<Z<2.41)}{P(Z>0)} = \frac{0.4920}{0.5} = 0.984$.
   (j) ≈ 0.95 (by the table), 0.9541653 (using software).
   (k) By the table: $Q_1 \approx -.67, Q_3 \approx .67, \text{IQR} \approx 1.34$. Using software: $Q_1 \approx -0.6744898, Q_3 \approx 0.6744898, \text{IQR} \approx 1.348980$
   (l) $P(Z < 2.13) = 0.9834$. If two values are randomly and independently selected, the probability they are both less than 2.13 is $0.9834^2 = 0.9671$.

10. (a) 0.9968 (the area to the right of -2.73 under the standard normal curve).
    (b) 0.0032 (the area to the right of 2.73 under the standard normal curve).
    (c) 1.17.
    (d) -1.17.
    (e) 1.96.
    (f) -0.71.
    (g) 0.00052. The probability a single value is greater than 2.0 is 0.0228. The probability two values, independently sampled, are both greater than 2.0 is $0.0228^2 = 0.00052$.

11. 0 (the area to the right of 3178890557384735892750948 7 under the standard normal curve is minuscule).

6.4.2 Standardizing Normally Distributed Random Variables

12. (a) $P(X < 105) = P(Z < \frac{105-120}{20}) - P(Z < -0.75) = 0.2266$.
    (b) $P(X > 87) = P(Z > \frac{87-120}{20}) - P(Z > -1.65) = 0.9505$.
    (c) $P(92 < X < 108) = P(\frac{92-120}{20} < Z < \frac{108-120}{20}) = P(-1.4 < Z < -0.6) = 0.2743 - 0.0808 = 0.1935$.
    (d) $P(82 < X < 87) = P(\frac{82-120}{20} < Z < \frac{87-120}{20}) = P(-1.9 < Z < -1.65) = 0.0495 - 0.0287 = 0.0208$.
    (e) 0.
    (f) 103.2. The 20th percentile of the standard normal distribution is approximately -0.84. The 20th percentile of the distribution of X is approximately $\mu + \sigma z = 120 + 20(-0.84) = 103.2$. (Using software, a more precise value is 103.1676.)
    (g) 136.8. The 80th percentile of the standard normal distribution is approximately 0.84. The 80th percentile of the distribution of X is approximately $\mu + \sigma z = 120 + 20 \cdot 0.84 = 136.8$. (Using software, a more precise value is 136.8324.)
    (h) $IQR = Q_3 - Q_1$. The first quartile of X is approximately $120 + 20(-0.67) = 106.6$, and the third quartile is approximately $120 + 20 \cdot 0.67 = 133.4$. The IQR is approximately $133.4 - 106.6 = 26.8$. (This was calculated using the table, but there is quite a bit of rounding error. Using software, a more precise value is 26.97959.)

13. Let X represent the weight (in kg) of a randomly selected newborn female African elephant that is born in captivity. Then $X \sim N(95.1, 13.7^2)$.

    (a) $P(X > 120.0) = P(Z > \frac{120.0-95.1}{13.7}) = P(Z > 1.818) = 0.035$.
    (b) $P(X < 100.0) = P(Z < \frac{100.0-95.1}{13.7}) = P(Z < 0.358) = 0.640$.
    (c) $P(90.0 < X < 110.0) = P(\frac{90.0-95.1}{13.7} < Z < \frac{110-95.1}{13.7}) = P(-0.372 < Z < 1.088) = 0.862 - 0.355 = 0.507$.
    (d) The 30th percentile of the standard normal distribution is approximately -0.52. The 30th percentile of newborn weights is approximately $95.1+13.7(-0.52) = 88.0$ kg. (There is a little rounding error here. Using software, a more precise value is 87.92.)
    (e) The 80th percentile of the standard normal distribution is approximately 0.84. The 80th percentile of newborn weights is approximately $95.1 + 13.7 \cdot 0.84 = 106.6$ kg.

14. (a) If X has a normal distribution with mean μ and standard deviation σ, then $\frac{X-\mu}{\sigma}$ has the standard normal distribution. *True.*
    (b) If Z has a standard normal distribution, then the mean and standard deviation of Z are equal. *False.* $\mu = 0, \sigma = 1$.

(c) The first quartile of the standard normal distribution is greater than 0. *False. The first quartile is approximately* -0.67.

(d) The first quartile of any normal distribution is greater than 0. *False. Any of the percentiles can be negative, depending on the parameters of the distribution.*

(e) If X has a normal distribution, then $P(X > \mu) = P(X < \mu)$. *True. Since the normal distribution is symmetric about* μ, $P(X > \mu) = P(X < \mu) = 0.5$.

6.5 Normal Quantile-Quantile Plots: Is the Data Approximately Normality Distributed?

6.5.1 Examples of Normal QQ Plots for Different Distributions

15. (a) Since the points on the plot fall close to a straight line, the population from which the sample is drawn is likely approximately normal.
    (b) Since the points on the plot fall close to a straight line, the population from which the sample is drawn is likely approximately normal.
    (c) This type of plot is typical of a distribution that is symmetric, but has shorter tails than the normal distribution (perhaps the uniform distribution, or something similar).
    (d) This type of plot is typical of a distribution that is skewed to the right.

16. The normal quantile-quantile plot would look like the one in Figure 6.1.

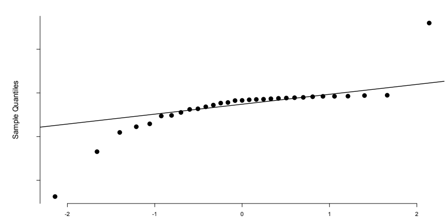


Figure 6.1: A normal quantile-quantile plot showing left skewness and one large outlier.

6.6 Other Important Continuous Probability Distributions

6.6.1 The χ^2 Distribution

6.6.2 The t Distribution

6.6.3 The F Distribution

6.7 Chapter Exercises

6.7.1 Basic Calculations

17. (a) 0.2. The height of the rectangle is $\frac{1}{50-0} = \frac{1}{50}$. $f(x) = \frac{1}{50}$ for $0 \leq x \leq 50$, and 0 otherwise. $P(X < 10) = 0.2$ (the area to the left of 10).
    (b) 0.8 (the area to the right of 10).
    (c) $P(12 < X < 23) = \frac{23-12}{50} = \frac{11}{50} = 0.22$.
    (d) $P(-5 < X < 5) = P(0 < X < 5)$ (since there is no area to the left of 0). $P(0 < X < 5) = (5-0) \times \frac{1}{50} = 0.1$.
    (e) $a = 10$ (since the area to the right of 10 is 0.8).
    (f) The 20th percentile is equal to 10 (since the area to the left of 10 is 0.2).
    (g) 25.

18. (a) 2. The area of a triangle is $\frac{1}{2} \times base \times height$. Since the area under the entire curve must equal 1, $\frac{1}{2} \times base \times height = 1 \implies \frac{1}{2} \times (1-0) \times height = 1 \implies height = 2$. So the highest point of the triangle is 2, which occurs at $x = 1$.
    (b) $f(0) = 0$ and $f(1) = 2$, which implies that $f(0.5) = 1$. (Since $f(x)$ is a straight line in this question.)
    (c) $f(x) = 2x$ for $0 < x < 1$ and 0 otherwise. (The curve is a line with slope = 2.)
    (d) 0.
    (e) 1. $P(X < 1.5)$ is the area to the left of 1.5. Since all of the area lies between 0 and 1, $P(X < 1.5) = 1$.
    (f) 0.25. $P(X < 0.5)$ is the area to the left of 0.5. At $x = 0.5$, $f(0.5) = 1$. The area to the left of 0.5 is a triangle with a base of $(0.5 - 0) = 0.5$ and a height of 1, so $P(X < 0.5) = \frac{1}{2} \times 0.5 \times 1 = 0.25$.
    (g) $P(X > 0.5) = 1 - P(X < 0.5) = 1 - 0.25 = 0.75$.
    (h) $\sqrt{0.5}$. We need to find the value a such that the area to the left of a is 0.5. The area to the left of a point a is $\frac{1}{2} \times (a-0) \times 2a$. Setting this equal to 0.5,

$$a^2 = 0.5 \implies a = \sqrt{0.5}.$$

6.7.2 Concepts

19. 1, since the area to the left of -458958643961346 under the standard normal curve is minuscule.

20. (a) The 95th percentile of a continuous random variable cannot be less than 0. *False.*
    (b) The mean of a continuous random variable cannot be negative. *False. Continuous random variables often have negative means.*
    (c) The mean of a continuous random variable can be greater than the third quartile. *True, but it would be unusual.*
    (d) The normal distribution is symmetric. *True.*
    (e) If X has a normal distribution, then the mean and median of X are equal. *True, since for symmetric distributions the mean and median are equal.*

6.7.3 Applications

21. Let X represent the total serum cholesterol (mmol/L) of a randomly selected Canadian male between 60 and 79 years of age. Then $X \sim N(4.7, 1.2^2)$.

    (a) $P(X \geq 6.2) = P(Z \geq \frac{6.2-4.7}{1.2}) = P(Z > 1.25) = 0.1056$.
    (b) 0.1056. (The proportion of men in this age group with a total serum cholesterol of at least 6.2 mmol/L is the same as the probability a randomly selected man in this age group has a level at least this high (which we found in 21a).)
    (c) 0.467. $P(X < 4.6) = P(X < \frac{4.6-4.7}{1.2}) = P(Z < -0.083) = 0.467$. (There will be a little rounding error if you use a table: $P(Z < -0.08) = 0.4681$.)
    (d) 0.428. $P(4.6 < X < 6.2) = P(\frac{4.6-4.7}{1.2} < Z < \frac{6.2-4.7}{1.2}) = P(-0.083 < Z < 1.25) = 0.894 - 0.467 = 0.428$. (There will be a little rounding error if you use a table: $P(Z < -0.08) = 0.4681$.)
    (e) 3.69 mmol/L. The 20th percentile of the standard normal distribution is approximately -0.84 and thus the 20th percentile of total serum cholesterol levels is approximately $\mu + \sigma z = 4.7 + 1.2(-0.84) = 3.69$ mmol/L.
    (f) 5.71 mmol/L. The 80th percentile of the standard normal distribution is approximately 0.84 and thus the 80th percentile of cholesterol levels is approximately $\mu + \sigma z = 4.7 + 1.2 \cdot 0.84 = 5.71$ mmol/L.

22. Let X represent the height (in cm) of a randomly selected Canadian female between 20 and 39 years of age. Then $X \sim N(163.3, 6.4^2)$.

(a) $P(X \geq 182.88) = P(Z \geq \frac{182.88-163.3}{6.4}) = P(Z \geq 3.059) = 0.0011$.
(b) $P(X > 170.0) = P(Z > \frac{170.0-163.3}{6.4}) = P(Z > 1.047) = 0.148$. (There will be a little rounding error if you are using a table: $P(Z > 1.05) = 0.1469$.)
(c) $P(X < 170.0) = P(Z < \frac{170.0-163.3}{6.4}) = P(Z < 1.047) = 0.852$. (There will be a little rounding error if you are using a table: $P(Z < 1.05) = 0.8525$.)
(d) $P(X > 150.0) = P(Z > \frac{150.0-163.3}{6.4}) = P(Z > -2.078) = 0.981$.
(e) $P(-1 < Z < 1) = 0.8413 - 0.1587 = 0.683$.
(f) 157.9 cm. The 20th percentile of the standard normal distribution is approximately -0.84. Thus the 20th percentile of female height is approximately $\mu + \sigma z = 163.3 + 6.4(-0.84) = 157.9$ cm. (Using software, a more precise value is 157.91.)

23. Let X represent the height (in cm) of a randomly selected Canadian male between 20 and 39 years of age. Then $X \sim N(177.7, 6.5^2)$.

(a) $P(X \geq 182.88) = P(Z \geq \frac{182.88-177.7}{6.5}) = P(Z \geq 0.797) = 0.213$. (There will be a little rounding error if you are using a table: $P(Z \geq 0.80) = 0.2119$.)
(b) $P(X > 170.0) = P(Z > \frac{170.0-177.7}{6.5}) = P(Z > -1.184) = 0.882$. (There will be a little rounding error if you are using a table: $P(Z > 1.18) = 0.881$.)
(c) $P(X < 170.0) = P(Z < \frac{170.0-177.7}{6.5}) = P(Z < -1.184) = 0.118$. This is 1 minus the probability found in part 23b.
(d) The 99th percentile of the standard normal distribution is approximately 2.33. The 99th percentile of the heights of Canadian males in this age group is approximately $\mu + \sigma z = 177.7 + 6.5 \times 2.33 = 192.8$ cm. (Using software, a more precise value is 192.82.)
(e) In 23b, we found the probability a randomly selected Canadian male in this age group is taller than 170.0 cm is 0.882. The probability that three randomly and independently selected males are *all* taller than 170.0 cm is $0.882^3 = 0.686$.
(f) In 23a, we found the probability a randomly selected Canadian male in this age group is at least 6 feet fall is 0.213. For two randomly selected males in this age group, the probability they are both at least 6 feet tall is $0.213^2 = 0.045$.
(g) The probability that a randomly selected Canadian male in this age group is at least 6 feet tall is 0.213 (found in 23a). If 5 males are randomly selected from this age group, the number that are at least 6 feet ball will have a binomial distribution with $n = 5$ and $p = 0.213$. If we let Y be the number that are at least 6 feet tall, then $P(Y = 2) = \binom{5}{2} 0.213^2(1 - 0.213)^3 = 0.221$.

24. Let X represent the test diameter (in mm) of a randomly selected mature male green sea urchin in this area. Then $X \sim N(39.7, 3.8^2)$.

(a) 0.290. $P(X > 40) = P(Z > \frac{40-37.9}{3.8}) = P(Z > 0.553) = 0.290$. (There will be a little rounding error if you are using a table: $P(Z > 0.55) = 0.2912$.)
(b) 0.019. $P(X < 30) = P(Z < \frac{30-37.9}{3.8}) = P(Z < -2.079) = 0.019$.

(c) 0.691. $P(30 < X < 40) = P(\frac{30-37.9}{3.8} < Z < \frac{40-37.9}{3.8}) = P(-2.079 < Z < 0.553) = 0.710 - 0.019 = 0.691$. (There will be a little rounding error if you are using a table: $P(Z < 0.55) = 0.7088$ and $P(Z < -2.08) = 0.0188$, so $P(-2.08 < Z < 0.55) = 0.690$.)

(d) 37.9 mm. (The normal distribution is symmetric, so the mean and median are equal.)

(e) 35.34 mm. The 25th percentile of the standard normal distribution is approximately -0.67. Thus the 25th percentile of test diameters is approximately $\mu + \sigma z = 37.9 + 3.8(-0.67) = 35.35$ mm. (This answer contains a little rounding error. Using software, we can find that the 25th percentile of the standard normal distribution is -0.67449, and the 25th percentile of test diameters is 35.34 mm.)

25. Let X represent the weight (in grams) of the cereal in a randomly selected bag of cereal of this type. Then $X \sim N(374.5, 2.8^2)$.

(a) 0.010. $P(X \geq 380) = P(Z \geq \frac{380-374.5}{2.8}) = P(Z \geq 1.964) = 0.025$.

(b) 0.010. $P(X < 368) = P(Z < \frac{368-374.5}{2.8}) = P(Z < -2.321) = 0.010$.

(c) 0.936. $P(370 < X < 380) = P(\frac{370-374.5}{2.8} < Z < \frac{380-374.5}{2.8}) = P(-1.607 < Z < 1.964) = 0.975 - 0.054 = 0.921$.

(d) 374.5 grams. (The normal distribution is a symmetric distribution, and the mean and median are equal in symmetric distributions.)

(e) 371.6 grams. The 15th percentile of the standard normal distribution is approximately -1.04. Thus the 15th percentile of cereal weight is approximately $\mu + \sigma z = 174.5 + 2.8(-1.04) = 371.6$ grams. (There is a little rounding error when finding a percentile of the standard normal distribution using the table. Using software, we can find that the 15th percentile of the standard normal distribution is -1.036433, and the 15th percentile of cereal weight is is 371.598 grams.)

26. Let X represent the amount of protein (in grams) of a randomly selected cup of this type of 2% milk. Then $X \sim N(8.0, 0.30^2)$.

(a) 0.631. $P(X < 8.10) = P(Z < \frac{8.1-8.0}{0.30}) = P(Z < 0.333) = 0.631$. (If you use a table, there will be a little rounding error: $P(Z < 0.33) = 0.6293$.)

(b) $P(X > 14.0) = P(Z > \frac{14.0-8.0}{0.30}) = P(Z > 20) \approx 0$.

(c) 0.125. $P(8.05 < X < 8.20) = P(\frac{8.05-8.0}{0.30} < Z < \frac{8.20-8.0}{0.30}) = P(0.167 < Z < 0.50) = 0.691 - 0.566 = 0.125$. (If you use a table, there will be a little rounding error: $P(Z < 0.17) = 0.5674$.)

(d) 8.70 grams. The 99th percentile of the standard normal distribution is approximately 2.33. Thus the 99th percentile of protein in a cup of 2% milk is approximately $\mu + \sigma z = 8.0 + 0.3 \cdot 2.33 = 8.70$ grams. (There is a little rounding error when finding a percentile of the standard normal distribution using the table. Using software, we can find that the 99th percentile of the standard normal distribution is 2.326348.)

(e) 8.20 grams. The 75th percentile of the standard normal distribution is approximately 0.67. Thus the 75th percentile of protein in a cup of 2% milk is approximately $\mu + \sigma z = 8.0 + 0.3 \cdot 0.67 = 8.20$ grams. (There is a little rounding error when finding a percentile of the standard normal distribution using the table. Using software, we can find that the 75th percentile of the standard normal distribution is 0.6744898.)

27. Let X represent the total amount of fat (in grams) in a randomly selected serving (85 grams) of this type of cooked ground beef. Then $X \sim N(10.2, 1.8^2)$.

(a) 0.996. $P(X < 15.0) = P(Z < \frac{15.0-10.2}{1.8}) = P(Z < 2.667) = 0.996$.

(b) 0.730. $P(8.0 < X < 12.0) = P(\frac{8.0-10.2}{1.8} < Z < \frac{12.0-10.2}{1.8}) = P(-1.222 < Z < 1.000) = 0.841 - 0.111 = 0.730$.

(c) 0.0003. For a single serving: $P(X > 14.0) = P(Z > \frac{14.0-10.2}{1.8}) = P(Z > 2.111) = 0.0174$. The probability that two randomly selected servings both have more than 14.0 grams of fat is $0.0174^2 = 0.0003$.

(d) 8.33 grams. The 15th percentile of the standard normal distribution is approximately -1.04. Thus the 15th percentile of fat in a serving of this type of beef is approximately $\mu + \sigma z = 10.2 + 1.8(-1.04) = 8.33$ grams. (There is a little rounding error when finding a percentile of the standard normal distribution using the table. Using software, we can find that the 15th percentile of the standard normal distribution is -1.036433, and the 15th percentile of fat content in a serving is 8.334 grams.)

(e) 13.16 grams. The 95th percentile of the standard normal distribution is approximately 1.645. Thus the 95th percentile of fat in a serving of this type of beef is approximately $\mu + \sigma z = 10.2 + 1.8 \cdot 1.645 = 13.16$ grams.

28. Let X represent the length (in mm) of the head of a randomly selected lizard of the species *P. versicolor*. Then $X \sim N(11.90, 0.41^2)$.

(a) 0.582. $P(11.0 < X < 12.0) = P(\frac{11.0-11.90}{0.41} < Z < \frac{12.0-11.90}{0.41}) = P(-2.195 < Z < 0.244) = 0.596 - 0.014 = 0.582$.

(b) 0.835. $P(X \geq 11.5) = P(Z \geq \frac{11.5-11.90}{0.41}) = P(Z \geq -0.976) = 0.835$. (If you use a table, there will be a little rounding error: $P(Z \geq -0.98) = 0.8365$.)

(c) The probability that a randomly selected adult female lizard of this species has a head length that is at least 22.5 mm is: $P(X \geq 22.5) = P(Z \geq \frac{22.5-11.90}{0.41}) = P(Z \geq 25.9) \approx 0$. Either the person caught an extremely unusual lizard, or this lizard is not of this species.

(d) 11.37 mm. The 10th percentile of the standard normal distribution is approximately -1.28. Thus the 10th percentile of head lengths in this type of lizard is approximately $\mu + \sigma z = 11.90 + 0.41(-1.28) = 11.38$ grams. (There is a little rounding error when finding a percentile of the standard normal distribution using the table. Using software, we can find that the 10th percentile of the standard normal distribution is -1.281552, and the 10th percentile of head lengths is 11.3745 mm.)

(e) 12.43 mm. The 90th percentile of the standard normal distribution is approximately 1.28. Thus the 90th percentile of head lengths in this type of lizard is approximately $\mu + \sigma z = 11.90 + 0.41 \cdot 1.28 = 12.42$ grams. (There is a little rounding error when finding a percentile of the standard normal distribution using the table. Using software, we can find that the 90th percentile of the standard normal distribution is 1.281552, and the 90th percentile of head lengths is 12.4254 mm.)

29. Let X represent the weight (in grams) of the heart of a randomly selected adult Caucasian male. Then $X \sim N(365, 71^2)$.

    (a) $P(X > 400) = P(Z > \frac{400-365}{71}) = P(Z > 0.493) = 0.311$. (There is a little rounding error if you use the table: $P(Z > 0.49) = 0.3121$.)
    (b) $P(X > 453.6) = P(Z > \frac{453.6-365}{71}) = P(Z > 1.248) = 0.106$.
    (c) $P(X < 300) = P(Z < \frac{300-365}{71}) = P(Z < -0.915) = 0.180$. (There is a little rounding error if you use the table: $P(Z < -0.92) = 0.1788$.)
    (d) $P(300 < X < 400) = P(\frac{300-365}{71} < Z < \frac{400-365}{71}) = P(-0.915 < Z < 0.493) = 0.689 - 0.180 = 0.509$.
    (e) The 90th percentile of the standard normal distribution is approximately 1.28. The 90th percentile of heart weight is approximately $365 + 71 \times 1.28 = 455.9$ grams. (This answer has a little rounding error. If we use software, we can find that the correct value to 2 decimal places is 455.99.)

30. Let X represent the weight (in kg) of a randomly selected newborn baby in the United States. Then $X \sim N(3.3, 0.5^2)$.

    (a) $P(X < 2.0) = P(Z < \frac{2.0-3.3}{0.5}) = P(Z < -2.60) = 0.005$.
    (b) $P(X < 4.0) = P(Z < \frac{4.0-3.3}{0.50}) = P(Z < 1.40) = 0.919$.
    (c) $P(2.0 < X < 4.0) = P(-2.60 < Z < 1.40) = 0.919 - 0.005 = 0.914$.
    (d) $P(X > 6.0) = P(Z > \frac{6.0-3.3}{0.50}) = P(Z > 5.4)$, which is very close to 0. (Using software, a more precise value is 3.3×10^{-8}. But it's not wise to count on the normal approximation to provide a great deal of precision this far out in the tail. In any event, the probability is near 0.)
    (e) The 90th percentile of the standard normal distribution is approximately 1.28, and thus the 90th percentile of full-term newborn baby weight is approximately $\mu + \sigma z = 3.3 + 0.50 \times 1.28 = 3.94$ kg.
    (f) The 25th percentile of the standard normal distribution is approximately -0.67, and thus the 90th percentile of full-term newborn baby weight is approximately $\mu + \sigma z = 3.3 + 0.50(-0.67) = 2.96$ kg.

6.7.4 Extra Practice Questions

31. Let X represent the lifetime (in days) of a randomly selected insect of this type. Then $X \sim N(60.2, 18.0^2)$.

    (a) $P(Z > \frac{100-60.2}{18.0}) = P(Z > 2.211) = 0.014$.
    (b) $P(Z > \frac{200-60.2}{18.0}) = P(Z > 7.767) \approx 0$. (Using software, a more precise value is 4.0×10^{-15}.)
    (c) $P(Z < \frac{50-60.2}{18.0}) = P(Z < -0.567) = 0.285$. (If you use the table, there will be a little rounding error: $P(Z < -0.57) = 0.2843$.)
    (d) $P(\frac{50-60.2}{18.0} < Z < \frac{70-60.2}{18.0}) = P(-0.567 < Z < -0.544) = 0.707 - 0.285 = 0.421$.
    (e) The 90th percentile of the standard normal distribution is approximately 1.28, and thus the 90th percentile of life lengths is approximately $\mu + \sigma z = 60.2 + 18 \times 1.28 = 83.2$ days. (This answer has a little rounding error. If we use software, we can find that the correct value is 83.27.)

32. Let X represent the purity level in a randomly selected oxygen container of this type. Then $X \sim N(99.50, 0.048^2)$.

    (a) $P(X < 99.43) = P(Z < \frac{99.43-99.50}{0.048}) = P(Z < -1.458) = 0.072$.
    (b) $P(Z > \frac{99.60-99.50}{0.048}) = P(Z > 2.083) = 0.019$.
    (c) The 10th percentile of the standard normal distribution is approximately -1.28, and thus the 10th percentile of purity levels is approximately $\mu + \sigma z = 99.50 + 0.048(-1.28) = 99.44$.
    (d) The 25th percentile (Q_1) of the standard normal distribution is approximately -0.67, and thus the 25th percentile of purity levels is approximately $\mu + \sigma z = 99.50 + 0.048(-0.67) = 99.468$. Q_3 is approximately $\mu + \sigma z = 99.50 + 0.048 \cdot 0.67 = 99.532$.
    (e) The probability that a single container has a purity level greater than 99.40% is $P(Z > \frac{99.40-99.50}{0.048}) = P(Z > -2.083) = 0.981$. The probability that 4 independently sampled containers *all* have purity levels greater than 99.40 is $0.981^4 = 0.93$.

33. Let X represent the weight (in grams) of a randomly selected adult pigeon in this city. Then $X \sim N(370, 28^2)$.

    (a) $P(X < 350) = P(Z < \frac{350-370}{28}) = P(Z < -0.714) = 0.238$. (If you use a table, there will be a little rounding error: $P(Z < -0.71) = 0.2389$.)
    (b) $P(X > 453.6) = P(Z > \frac{453.6-370}{28}) = P(Z > 2.986) = 0.0014$.
    (c) $P(350 < X < 400) = P(\frac{350-370}{28} < Z < \frac{400-370}{28}) = P(-0.714 < Z < 1.071) = 0.858 - 0.238 = 0.620$.
    (d) $P(X > 6000) = P(Z > \frac{6000-370}{28}) = P(Z > 201.07) \approx 0$. (Under the assumed

model, this probability is non-zero, but very very close to 0. For all intents and purposes there is no chance of observing this event.)

(e) The 5th percentile of the standard normal distribution is approximately -1.645. The 5th percentile of pigeon weights is approximately $\mu+\sigma z = 370+28(-1.645) = 323.94$ grams.

34. Let X represent the cheese demand (in kg) on a randomly selected day. Then $X \sim N(50, 36)$. Note that the variance is given ($\sigma^2 = 36$), and the standard deviation is therefore $\sigma = \sqrt{36} = 6$ kg.

    (a) $P(X > 60) = P(Z > \frac{60-50}{\sqrt{36}}) = P(Z > 1.667) = 0.048$.
    (b) $P(X < 45) = P(Z < \frac{45-50}{\sqrt{36}}) = P(Z < -0.833) = 0.202$. (If you use a table, there will be a little rounding error: $P(Z < -0.83) = 0.2032$.)
    (c) The 80th percentile of the standard normal distribution is approximately 0.84 and thus the 80th percentile of cheese demand is approximately $\mu + \sigma z = 50 + \sqrt{36} \times 0.84 = 55.04$ kg. (There is a little rounding error here. If we use software we can find the correct value of 55.05 kg.)
    (d) This is equivalent to asking for the 99th percentile of cheese demand. The 99th percentile of the standard normal distribution is approximately 2.33 and thus the 99th percentile of cheese demand is approximately $\mu + \sigma z = 50 + \sqrt{36} \times 2.33 = 63.98$ kg. (There is a little rounding error here. If we use software we can find the correct value of 63.96 kg.)

35. Let X represent the diameter (in cm) of a randomly selected shaft. Then $X \sim N(4.10, 0.05^2)$.

    (a) $P(X > 4.23) = P(Z > \frac{4.23-4.10}{0.05}) = P(Z > 2.6) = 0.0047$.
    (b) $P(X < 4.00) = P(Z < \frac{4.00-4.10}{0.05}) = P(Z < -2.0) = 0.0228$.
    (c) The 10th percentile of the standard normal distribution is approximately -1.28, and thus the 10th percentile of diameters is approximately $\mu + \sigma z = 4.10 + 0.05(-1.28) = 4.036$ cm.
    (d) The 90th percentile of the standard normal distribution is approximately 1.28, and thus the 90th percentile of diameters is approximately $\mu + \sigma z = 4.10 + 0.05 \cdot 1.28 = 4.164$ cm.
    (e) $P(\text{Shaft must be scrapped}) = P(X < 4.00) + P(X > 4.20) = P(Z < \frac{4.00-4.10}{0.05}) + P(Z > \frac{4.20-4.10}{0.05}) = P(Z < -2.0) + P(Z > 2.0) = 0.0228 + 0.0228 = 0.0456$.

Chapter 7

Sampling Distributions

J.B.'s strongly suggested exercises: 1, 6, 7, 8, 9, 10, 20, 21, 22, 23, 24, 25

7.1 Introduction

1. (a) We view statistics as random variables that have probability distributions. *True.*
   (b) We view parameters as random variables that have probability distributions. *False. A parameter's value is not usually known, but it is a fixed value that does not change from sample to sample.*
   (c) The sampling distribution of a statistic is the probability distribution of the statistic. *True. "Sampling distribution" is another name for the probability distribution of a statistic.*
   (d) In repeated sampling, the value of a statistic will vary about the parameter it estimates. *True.*
   (e) In repeated sampling, the value of a parameter will vary about the statistic that estimates it. *False. A parameter's value is not usually known, but it is a fixed value that does not change from sample to sample.*

2. (a) $\mu = 14$.
   (b) $\binom{4}{2} = 6$.

(c)

Sample values	Sample mean	Probability of occurring
12, 13	12.5	1/6
12, 15	13.5	1/6
12, 16	14.0	1/6
13, 15	14.0	1/6
13, 16	14.5	1/6
15, 16	15.5	1/6

(d)

$\bar{x}$	12.5	13.5	14.0	14.5	15.5
$p(\bar{x})$	1/6	1/6	2/6	1/6	1/6

(e)

$$\mu_{\bar{X}} = E(\bar{X}) = \sum \bar{x}p(\bar{x}) = 12.5 \cdot \frac{1}{6} + 13.5 \cdot \frac{1}{6} + 14.0 \cdot \frac{2}{6} + 14.5 \cdot \frac{1}{6} + 15.5 \cdot \frac{1}{6} = 14$$

Note that $E(\bar{X}) = \mu$ (on average, the sample mean equals the population mean).

3. (a)

Sample values	Sample variance	Probability of occurring
12, 13	0.5	1/6
12, 15	4.5	1/6
12, 16	8.0	1/6
13, 15	2.0	1/6
13, 16	4.5	1/6
15, 16	0.5	1/6

(b)

s^2	0.5	2.0	4.5	8.0
$p(s^2)$	2/6	1/6	2/6	1/6

(c) $\frac{2}{6} = \frac{1}{3}$.

4. (a) Recall that the probability distribution of a discrete random variable is a listing of all possible values of that random variable and their probabilities of occurring. So the probability distribution of X is:

x	0	1
$p(x)$	0.5	0.5

This is a Bernoulli distribution with $p = 0.5$. (We'll get heads half the time.)

(b) $\mu_X = E(X) = \sum xp(x) = 0 \cdot 0.50 + 1 \cdot 0.50 = 0.50$.

(c) $\sigma_X^2 = \sum (x - \mu)^2 p(x) = (0 - 0.5)^2 \cdot 0.5 + (1 - 0.5)^2 \cdot 0.5 = 0.25$.

(d) The following table gives the 4 possible outcomes, the corresponding values of $\bar{X}$, and their probabilities of occurring.

The probability distribution of the sample mean $\bar{X}$ is thus:

$\bar{x}$	0	0.5	1
$p(\bar{x})$	0.25	0.5	0.25

Outcome	TT	TH	HT	HH
Value of $\bar{X}$	0	0.5	0.5	1
Probability	0.25	0.25	0.25	0.25

This is the sampling distribution of the average number of heads per toss ($\bar{X}$) when a fair coin is tossed twice. It was not necessary to actually toss the coin to derive the sampling distribution.

(e) The mean of the sampling distribution of $\bar{X}$ is represented by $\mu_{\bar{X}}$. We can use the formula for the mean of a discrete random variable to find the mean of $\bar{X}$:

$$\mu_{\bar{X}} = E(\bar{X}) = \sum \bar{x}p(\bar{x}) = 0 \cdot 0.25 + 0.5 \cdot 0.5 + 1 \cdot 0.25 = 0.50$$

Note that the mean of the probability distribution of the sample mean is equal to mean of the probability distribution for a single toss: $\mu_{\bar{X}} = \mu_X = \frac{1}{2}$.

(f) The variance of the sampling distribution of $\bar{X}$ is represented by $\sigma^2_{\bar{X}}$. We can use the formula for the variance of a discrete random variable to find the variance of $\bar{X}$:

$$\sigma^2_{\bar{X}} = \sum(\bar{x}-\mu_{\bar{X}})^x p(\bar{x}) = (0 = 0.5)^2 \cdot 0.25+(0.5-0.5)^2 \cdot 0.5+(1-0.5)^2 \cdot 0.25 = 0.125$$

Note that the variance of $\bar{X}$ is the variance of X divided by the number of tosses: ($\sigma^2_{\bar{X}} = \frac{\sigma^2_X}{2} = \frac{0.25}{2} = 0.125$).

5. (a) Let X represent the number of winning cups. Then X has a binomial distribution with $n = 2$ and $p = \frac{1}{6}$. The distribution of X in table form is:

x	0	1	2
$p(x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

(b) The proportion of winning cups is the number of winning cups divided by the number of cups purchased ($\hat{p} = \frac{X}{n}$). So the sample proportion takes on 3 possible values: $\hat{p} = \frac{0}{2} = 0$, or $\hat{p} = \frac{1}{2} = 0.5$, or $\hat{p} = \frac{2}{2} = 1$. The sampling distribution of the sample proportion is:

$\hat{p}$	0	0.5	1
Probability	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

7.2 The Sampling Distribution of the Sample Mean

6. (a) $\mu_{\bar{X}} = \mu = 10$.

(b) $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \frac{\sqrt{625}}{\sqrt{16}} = 6.25$.

(c) $\bar{X} \sim N(10, \frac{625}{16})$.

(d) $P(\bar{X} > 30) = P(Z > \frac{30-\mu}{\sigma/\sqrt{n}}) = P(Z > \frac{30-10}{25/\sqrt{16}}) = P(Z > 3.2) = 0.0007.$

(e) $P(0 < \bar{X} < 20) = P(\frac{0-10}{25/\sqrt{16}} < Z < \frac{20-10}{25/\sqrt{16}}) = P(-1.6 < Z < 1.6) = 0.8904.$

7. Plot B is the sampling distribution of the sample mean. In this situation the sampling distribution of $\bar{X}$ will be approximately normal with a mean of 20 and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{400}} = 1$. Although A, B, and D all have means of 20, plot B is the only one with a standard deviation that appears to be close to 1.

7.3 The Central Limit Theorem

8. The central limit theorem tells us that sums and means of random variables have distributions that tend toward the normal distribution as the number of observations increases, regardless of the distribution from which we are sampling. This implies that for large sample sizes, the distribution of the sample mean will be approximately normal, even when we are sampling from a non-normal population.

   The central limit theorem is very important in the world of statistics. Because of the CLT, we can often make probability calculations and use statistical inference procedures based on the normal distribution, even when the underlying population is not normally distributed.

9. False. μ is a parameter, and as such it does not have a sampling distribution. (We view μ as a fixed, unchanging value.)

10. (a) The standard deviation of the sampling distribution of the sample mean depends on the value of μ. *False.* $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ *depends only on* σ *and* n.

    (b) We cannot possibly determine any characteristics of a statistic's sampling distribution without repeatedly sampling from the population. *False. We can often mathematically determine characteristics of the sampling distribution. For example, we have shown that* $E(\bar{X}) = \mu$.

    (c) The sampling distribution of $\bar{X}$ is always at least approximately normal for large sample sizes, and is sometimes approximately normal for small sample sizes. *True. The distribution of* $\bar{X}$ *is normal for any sample size if we are sampling from a normally distributed population. If we are sampling from a distribution that is approximately normal, then the distribution of* $\bar{X}$ *will be approximately normal for any sample size.*

    (d) If the sample size is quadrupled, then standard deviation of the sampling distribution of the sample mean would decrease by a factor of 2. *True.* $\frac{\sigma}{\sqrt{n}}$ *decreases by a factor of* $\sqrt{4} = 2$ *if* $n = 4$.

7.4 Some Terminology Regarding Sampling Distributions

7.4.1 Standard Errors

11. The standard error of a statistic is the estimate of the standard deviation of that statistic's sampling distribution. (It is a measure of the variability of the statistic in repeated sampling.)

    (Note: Some sources define standard error as the true standard deviation of the statistic's sampling distribution, and the *estimated* standard error as the estimate. But here we reserve the term *standard error* for the estimated standard deviation of a statistic's sampling distribution.)

    We view parameters as fixed, usually unknown quantities. As such, parameters do not have sampling distributions or standard errors.

12. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{13.7}{\sqrt{6}} = 5.593$.

13. (a) Increasing the sample size decreases $SE(\bar{X})$.
    (b) An increase in the sample variance increases $SE(\bar{X})$.
    (c) $SE(\bar{X})$ is affected by the variability and the sample size, abut not by the value of $\bar{X}$. The sample mean has no effect on $SE(\bar{X})$.

7.4.2 Unbiased Estimators

14. A statistic is an unbiased estimator of a parameter if its expected value is equal to that parameter. (In other words, a statistic is an unbiased estimator if, on average, the statistic equals the parameter it estimates.)

15. Yes, since $E(\bar{X}) = \mu$.

7.5 Chapter Exercises

7.6 Basic Calculations

7.7 Concepts

16. (a) The letter n represents the number of samples that are drawn in repeated sampling, and not the number of observations in each sample. *False. n represents the number of observations in the sample.*

    (b) In practice, we usually take a large number of samples so that we can accurately estimate the sampling distribution of $\bar{X}$. *False. In practice, we draw only a single sample. (We may draw a sample with a large number of observations, but we draw only a single sample.) We use mathematical techniques to derive characteristics of the sampling distribution of $\bar{X}$.*

    (c) The standard deviation of the population from which we are sampling decreases as the sample size increases. *False. The standard deviation of the population is what it is, and drawing more observations won't change that.*

    (d) The standard deviation of the sampling distribution of $\bar{X}$ decreases as the sample size increases. *True.* $\frac{\sigma}{\sqrt{n}}$ *decreases as n increases.*

17. (a) The probability distribution of X is:

x	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

    (b) $E(X) = \sum xp(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$

    (c) $\sigma^2 = \sum (x-\mu)^2 p(x) = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} = \frac{35}{12}.$

    (d) To find the sampling distribution of $\bar{X}$, we can list all 36 possible outcomes and calculate $\bar{X}$ for each possibility. The results are given in the following table.

Roll 1 / Roll 2	1	2	3	4	5	6
1	1.0	1.5	2.0	2.5	3.0	3.5
2	1.5	2.0	2.5	3.0	3.5	4.0
3	2.0	2.5	3.0	3.5	4.0	4.5
4	2.5	3.0	3.5	4.0	4.5	5.0
5	3.0	3.5	4.0	4.5	5.0	5.5
6	3.5	4.0	4.5	5.0	5.5	6.0

    The 36 possible outcomes are equally likely, so the probability of obtaining each value of $\bar{X}$ is just the proportion of times that value of $\bar{X}$ occurs. The probability distribution of $\bar{X}$ is:

$\bar{x}$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
$p(\bar{x})$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

(e) The quick way is to realize that $\mu_{\bar{X}} = \mu = 3.5$ (found in 17b). The slower way would be to use the formula for the expectation of a discrete random variable: $E(\bar{X}) = \sum \bar{x}p(\bar{x})$.

(f) The quick way is to realize that $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{35/12}}{\sqrt{2}} \approx 1.2076$. (In 17c we found that $\sigma^2 = \frac{35}{12}$.) The slower way is to use the formula for the variance of a discrete random variable: $\sigma^2_{\bar{X}} = \sum(\bar{x} - \mu)^2 p(\bar{x})$.

18. (a) $\mu_{\bar{X}} = \mu = 3.5$.

(b) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{35/12}}{\sqrt{2000}} \approx 0.03818813$.

(c) The central limit theorem tells us that $\bar{X}$ will be approximately normally distributed (with a mean of 3.5 and a standard deviation of approximately 0.038188).

(d) $P(\bar{X} < 3.5190) = P(Z < \frac{3.5190-3.5}{0.03818813}) = P(Z < 0.4975368) = 0.691$.

19. (a) If we are sampling from a normally distributed population, then the sampling distribution of $\bar{X}$ is normal for any sample size. *True*

(b) The standard deviation of the sampling distribution of the sample mean increases as the sample size increases. *False.* $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ *decreases as the sample size increases.*

(c) The central limit theorem states that the sample mean equals the population mean, as long as the sample size is large. *False.* $E(\bar{X}) = \mu$ *(for any sample size) but there will always be some variability (the value of* $\bar{X}$ *will not be exactly equal to* μ*).*

(d) If the sampling distribution of $\bar{X}$ is approximately normal, then the population from which we are sampling must be approximately normal. *False. For large sample sizes the sampling distribution of* $\bar{X}$ *is always approximately normal, regardless of the shape of the population from which we are sampling, and thus an approximately normal sampling distribution of* $\bar{X}$ *does not say much about the sampled population.*

(e) All else being equal, the mean of the sampling distribution of $\bar{X}$ decreases as the sample size increases. *False. The mean of the sampling distribution of* $\bar{X}$ *is* μ*, regardless of the sample size.*

20. (a) If we draw a very large sample from any population and plot a histogram of the observations, the shape of the histogram will be approximately normal. *False. A histogram of a large number of observations sampled from a distribution will*

look like the distribution from which we are sampling.

(b) In practice, we usually know the true standard deviation of the sampling distribution of $\bar{X}$. *False. The population standard deviation σ is almost always unknown, and so the true standard deviation of the sampling distribution of $\bar{X}$ ($\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$) is almost always unknown.*

(c) In practice, we usually know the true value of μ. *False. In practical problems, the parameter μ is almost always unknown.*

(d) The sample mean is an unbiased estimator of the population mean. *True.* $E(\bar{X}) = \mu$.

(e) If we were to repeatedly sample from a population, then the distribution of the sample mean would become approximately normal as the number of samples increases, as long as the sample size of each sample stays constant. *False. The distribution of $\bar{X}$ becomes approximately normal as the sample size n increases (in other words, as the number of values used to calculate the mean increases). Repeatedly sampling from a population doesn't change the sampling distribution of $\bar{X}$.*

21. (a) Statistics have sampling distributions. *True.*

(b) The value of a parameter does not vary from sample to sample. *True.*

(c) The value of a statistic does not vary from sample to sample. *False.*

(d) All else being equal, the standard deviation of the sampling distribution of the sample mean will be smaller for $n = 10$ than for $n = 40$. *False.* $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ *decreases for increasing sample size.*

(e) The sampling distribution of μ is usually approximately normal for $n > 30$. *False. We view μ as a fixed, unknown quantity. As such, it does not vary from sample to sample and does not have a sampling distribution.*

7.8 Applications

22. Let X represent the amount of riboflavin (in mg) in a randomly selected large white egg. Then $X \sim N(0.25, 0.02^2)$.

(a) $P(X < 0.20) = P(Z < \frac{0.20-0.25}{0.02}) = P(Z < -2.5) = 0.0062$.

(b) $\bar{X} \sim N(0.25, \frac{0.02^2}{30})$. (The mean amount of riboflavin in 30 large eggs is approximately normally distributed with a mean of $\mu_{\bar{X}} = 0.25$ mg and a variance of $\sigma^2_{\bar{X}} = \frac{0.02^2}{30} = 0.000013$ mg^2.)

(c) $P(0.240 < \bar{X} < 0.260) = P(\frac{0.240-0.25}{0.02/\sqrt{30}} < Z < \frac{0.260-0.25}{0.02/\sqrt{30}}) = P(-2.739 < Z < 2.739) = 0.994$.

(d) $P(0.250 < \bar{X} < 0.255) = P(\frac{0.250-0.25}{0.02/\sqrt{30}} < Z < \frac{0.255-0.25}{0.02/\sqrt{30}}) = P(0 < Z < 1.369) = 0.915 - 0.5 = 0.415$.

(e) 0.2469 mg. The 20th percentile of the standard normal distribution is approximately -0.84. The 20th percentile of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}} z = 0.25 + \frac{0.02}{\sqrt{30}}(-0.84) = 0.2469$ mg.

23. Let X represent the amount of protein (in grams) of a randomly selected breakfast sandwich of this type. Then $X \sim N(17.0, 0.8^2)$.

(a) $P(X < 16.0) = P(Z < \frac{16.0-17.0}{0.80}) = P(Z < -1.25) = 0.106$.
(b) $T \sim N(34.0, 1.28)$. Let $T = X_1 + X_2$ represent the total amount of protein in 2 randomly selected sandwiches. Then $E(T) = 17.0 + 17.0 = 34.0$ grams, and $Var(T) = Var(X_1) + Var(X_2) = 0.8^2 + 0.8^2 = 1.28$ grams^2. Since X_1 and X_2 are normal, their sum is normal.
(c) $\bar{X} \sim N(17.0, \frac{0.80^2}{2})$. (The mean amount of protein in 2 sandwiches is approximately normally distributed with a mean of $\mu_{\bar{X}} = 17.0$ g and a variance of $\sigma^2_{\bar{X}} = \frac{0.80^2}{2} = 0.32$ g^2.)
(d) 0.0385 There are 2 approaches that lead to the same result:
   i. We know from 23b that $T \sim N(34.0, 1.28)$ (where T represents the total amount of protein in 2 sandwiches). So $P(T < 32.0) = P(Z < \frac{32.0-34.0}{\sqrt{1.28}}) = P(Z < -1.768) = 0.0385$.
   ii. Alternatively, we could note that the total amount of protein will be less than 32.0 grams if the mean amount in the 2 sandwiches is less than 16.0 grams. $P(\bar{X} < 16.0) = P(Z < \frac{16.0-17.0}{0.80/\sqrt{2}}) = P(Z < -1.768) = 0.0385$.
(e) $P(\bar{X} \geq 16.5) = P(Z \geq \frac{16.5-17.0}{0.80/\sqrt{2}}) = P(Z \geq -0.884) = 0.812$.
(f) 18.109 grams. The 97.5th percentile of the standard normal distribution is approximately 1.96. The 97.5th percentile of the distribution of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}} z = 17.0 + \frac{0.80}{\sqrt{2}} \cdot 1.96 = 18.109$ grams.

24. Let X represent the length (in mm) of the right ear of a randomly selected Caucasian Italian man in this age group. Then $X \sim N(62, 4^2)$.

(a) $P(X \geq 60) = P(Z \geq \frac{60-62}{4}) = P(Z \geq -0.5) = 0.691$.
(b) 67.13 mm. (There is a little rounding error if you use a table. The 90th percentile of the standard normal distribution is approximately 1.28. The 90th percentile of right ear lengths is approximately $\mu + \sigma z = 62 + 4 \cdot 1.28 = 67.12$ mm.)
(c) $\bar{X} \sim N(62, 0.16)$. (The mean ear length is approximately normally distributed with a mean of $\mu_{\bar{X}} = 62$ mm and a variance of $\sigma^2_{\bar{X}} = \frac{4^2}{100} = 0.16$ mm^2.)
(d) $P(\bar{X} > 61.8) = P(Z > \frac{61.8-62}{4/\sqrt{100}}) = P(Z > -0.5) = 0.691$.
(e) 61.49 mm. The 10th percentile of the standard normal distribution is approximately -1.28. The 10th percentile of the distribution of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}} z = 62 + \frac{4}{\sqrt{100}}(-1.28) = 61.49$ mm.

25. Let X represent the length (in mm) of the carapace of a randomly selected adult male crayfish in this lake. Then $X \sim N(48.0, 2.5^2)$.

    (a) $P(X < 50.0) = P(Z < \frac{50.0-48.0}{2.5}) = P(Z < 0.8) = 0.788$.
    (b) $\bar{X} \sim N(48.0, 0.625)$. (The mean carapace length is approximately normally distributed with a mean of $\mu_{\bar{X}} = 48.0$ mm and a variance of $\sigma^2_{\bar{X}} = \frac{2.5^2}{10} = 0.625$ mm^2.)
    (c) $P(\bar{X} < 50.0) = P(Z < \frac{50.0-48.0}{2.5/\sqrt{10}}) = P(Z < 2.530) = 0.9943$.
    (d) $P(46.0 < \bar{X} < 50.0) = P(\frac{46.0-48.0}{2.5/\sqrt{10}} < Z < \frac{50.0-48.0}{2.5/\sqrt{10}}) = P(-2.530 < Z < 2.530) = 0.9886$.
    (e) 49.30 mm. The 95th percentile of the standard normal distribution is approximately 1.645. The 95th percentile of the distribution of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}}z = 48.0 + \frac{2.5}{\sqrt{10}} \cdot 1.645 = 49.30$ mm.

26. Let X represent the weight (in kg) of a randomly selected newborn female African elephant that is born in captivity. Then $X \sim N(95.1, 13.7^2)$.

    (a) 0.360. $P(X > 100.0) = P(Z > \frac{100.0-95.1}{13.7}) = P(Z > 0.358) = 0.360$. (There will be a little rounding error if you use a table: $P(Z > 0.36) = 0.3594$.)
    (b) $\bar{X} \sim N(95.1, \frac{13.7^2}{3})$. (The sampling distribution of the mean weight of 3 randomly selected newborn female elephants is approximately normal with a mean of 95.1 kg and a variance of $\frac{13.7^2}{3} \approx 62.6$ kg^2.
    (c) 0.268. $P(\bar{X} > 100.0) = P(Z > \frac{100.0-95.1}{13.7/\sqrt{3}}) = P(Z > 0.619) = 0.268$.
    (d) 0.196. $P(98.0 < \bar{X} < 102.0) = P(\frac{98.0-95.1}{13.7/\sqrt{10}} < Z < \frac{102.0-95.1}{13.7/\sqrt{10}}) = P(0.669 < Z < 1.593) = 0.944 - 0.748 = 0.196$.
    (e) 87.9 kg. Using a table, we would find that the 30th percentile of the standard normal distribution is approximately -0.52. The 30th percentile of birth weight is approximately $\mu + \sigma z = 95.1 + 13.7(-0.52) = 87.98$ kg. (But there is a fair bit of rounding error if you use a table. Using software, we can find that the 30th percentile of the standard normal distribution is -0.5244 and the 30th percentile of birth weights is 87.915 kg.)
    (f) 92.83 kg. Using a table, we would find that the 30th percentile of the standard normal distribution is approximately -0.52. The 30th percentile of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}}z = 95.1 + \frac{13.7}{\sqrt{10}}(-0.52) = 92.847$ kg. (But there is a fair bit of rounding error if you use a table. Using software, we can find that the 30th percentile of the standard normal distribution is -0.5244 and the 30th percentile of the sampling distribution of mean birth weight is 92.828 kg.)

27. Let X represent the weight (in grams) of the heart of a randomly selected adult Caucasian male. Then $X \sim N(365, 71^2)$.

    (a) 456.0 grams.

The 90th percentile of the standard normal distribution is approximately 1.28. The 90th percentile of the distribution of heart weight is approximately $\mu + \sigma z = 365 + 71 \cdot 1.28 = 455.88$. (But there is a fair bit of rounding error if you use a table. Using software, we can find that the 90th percentile of the standard normal distribution is 1.281552 and the 90th percentile of the sampling distribution of heart weight is 455.99 g.)

(b) $\bar{X} \sim N(365, \frac{71^2}{9})$. (The sample mean is approximately normally distributed with a mean of 365 grams and a variance of $\frac{71^2}{9} = 560$ grams^2.)

(c) 0.737. $P(\bar{X} > 350.0) = P(Z > \frac{350.0-365.0}{71.0/\sqrt{9}}) = P(Z > -0.634) = 0.737$. (There will be a little rounding error if you use a table: $P(Z > -0.63) = 0.7357$.)

(d) 0.591. $P(340.0 < \bar{X} < 380.0) = P(\frac{340.0-365.0}{71.0/\sqrt{9}} < Z < \frac{380.0-365.0}{71.0/\sqrt{9}}) = P(-1.056 < Z < 0.634) = 0.591$.

(e) 395.3 grams. The 90th percentile of the standard normal distribution is approximately 1.28. The 90th percentile of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}} z = 365 + \frac{71.0}{\sqrt{9}} \cdot 1.28 = 395.29$ grams. (But there is a fair bit of rounding error if you use a table. Using software, we can find that the 90th percentile of the standard normal distribution is 1.281552 and the 90th percentile of the sampling distribution of the mean heart weight is 395.33 grams.)

(f) 0.664. There are 2 approaches that lead to the same result:

i. The sum of the weights of 4 randomly selected adult Caucasian male hearts is approximately normally distributed with a mean of $4\mu = 4 \times 365 = 1460$ grams and a variance of $4\sigma^2 = 4 \times 71^2 = 20164$ grams^2. $P(\sum_{i=1}^{4} X_i > 1400) = P(Z > \frac{1400-1460}{\sqrt{20164}}) = P(Z > -0.423) = 0.664$.

ii. Alternatively, we could note that the sum will be greater than 1400 if the mean of the 4 observations is greater than $\frac{1400}{4} = 350$. $P(\bar{X} > 350) = P(Z > \frac{350-365}{71/\sqrt{4}}) = P(Z > -0.423) = 0.664$.

28. Let X represent the height (in cm) of a randomly selected Canadian female between 20 and 39 years of age. Then $X \sim N(163.3, 6.4^2)$.

(a) $P(X \geq 170) = P(Z \geq \frac{170.0-163.3}{6.4}) = P(Z \geq 1.047) = 0.148$. (There will be a little rounding error if you use a table: $P(Z \geq 1.05) = 0.1469$.)

(b) $\bar{X} \sim N(163.3, 0.4096)$. (The sample mean is approximately normally distributed with a mean of 163.3 cm and a variance of $\frac{\sigma^2}{n} = \frac{6.4^2}{100} = 0.4096$ cm^2.)

(c) 0.996. $P(\bar{X} < 165.0) = P(Z < \frac{165.0-163.3}{6.4/\sqrt{100}}) = P(Z < 2.656) = 0.996$.

(d) $P(163.0 < \bar{X} < 165.0) = P(\frac{163.0-163.3}{6.4/\sqrt{100}} < Z < \frac{165.0-163.3}{6.4/\sqrt{100}}) = P(-0.469 < Z < 2.656) = 0.996 - 0.320 = 0.676$.

(e) The 2.5th percentile of the standard normal distribution is approximately -1.96. The 2.5th percentile of the distribution of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}} z = 163.3 + \frac{6.4}{10}(-1.96) = 162.05$ cm.

The 97.5th percentile of the standard normal distribution is approximately

1.96. The 97.5th percentile of the distribution of $\bar{X}$ is approximately $\mu + \frac{\sigma}{\sqrt{n}} z = 163.3 + \frac{6.4}{10} \cdot 1.96 = 164.55$ cm.

7.8.1 Extra Practice Questions

29. (a) $\mu_{\bar{X}} = \mu = 71$.
    (b) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{100}}{\sqrt{5}} = 4.472$.
    (c) $\bar{X} \sim N(71, 20)$. (The sample mean is approximately normal with a mean of 71 and a variance of $\frac{\sigma^2}{n} = \frac{100}{5} = 20$.)
    (d) $P(X > 75) = P(Z > \frac{75-71}{10}) = P(Z > 0.4) = 0.345$.
    (e) $P(\bar{X} > 75) = P(Z > \frac{75-71}{10/\sqrt{5}}) = P(Z > 0.894) = 0.186$. (There will be a little rounding error if you use a table: $P(Z > 0.89) = 0.1867$).
    (f) From 29d, the probability a randomly selected student scored higher than 75 is 0.345. The probability all 5 randomly selected students scored higher than 75 is approximately $0.345^5 = 0.005$. (Note that for this calculation to be reasonable, we require a very large class size.)

30. Let X represent the diameter of a randomly selected shaft produced by this process. Then $X \sim N(4.10, 0.05^2)$.

    (a) $P(X < 4.23) = P(Z < \frac{4.23-4.10}{0.05}) = P(Z < 2.6) = 0.9953$.
    (b) $\bar{X} \sim N(4.10, \frac{0.05^2}{16})$. (The sample mean is approximately normally distributed with a mean of 4.10 cm and variance of $\frac{0.05^2}{16} = 0.00015625$ cm^2.)
    (c) $\bar{X} \sim N(4.10, \frac{0.05^2}{100})$. (The sample mean is approximately normally distributed with a mean of 4.10 cm and variance of $\frac{0.05^2}{100} = 0.000025$ cm^2.)
    (d) $P(4.05 < \bar{X} < 4.15) = P(\frac{4.05-4.10}{0.05/\sqrt{5}} < Z < \frac{4.15-4.10}{0.05/\sqrt{5}}) = P(-2.236 < Z < 2.236) = 0.975$.
    (e) $P(\bar{X} \geq 4.15) = P(Z \geq \frac{4.15-4.10}{0.05/\sqrt{5}}) = P(Z \geq 2.236) = 0.013$.
    (f) 4.114 cm. The 90th percentile of the standard normal distribution is approximately 1.28. The 90th percentile of $\bar{X}$ is approximately $\mu + \frac{\sigma}{n} z = 4.10 + \frac{0.05}{\sqrt{20}} \times 1.28 = 4.114$ cm.

31. Let X represent the oxygen purity level of a randomly selected container produced by this company. If the company's claim is true, $X \sim N(99.50, 0.048^2)$.

    (a) $P(X < 99.43) = P(Z < \frac{99.43-99.50}{0.048}) = P(Z < -1.458) = 0.072$.
    (b) $P(\bar{X} < 99.45) = P(Z < \frac{99.45-99.50}{0.048/\sqrt{10}}) = P(Z < -3.294) = 0.0005$.
    (c) $P(99.45 < \bar{X} < 99.55) = P(\frac{99.45-99.50}{0.048/\sqrt{10}} < Z < \frac{99.55-99.50}{0.048/\sqrt{10}}) = P(-3.294 < Z < 3.294) = 0.9990$.

(d) If it is reasonable to view this sample as a random sample of containers from this company, and the company's claim is true, the probability of observing a mean purity level this low or lower is $P(\bar{X} \leq 99.00) = P(Z \leq \frac{99.00-99.50}{0.048/\sqrt{25}}) = P(Z < -52)$. This probability is very, very near 0. There is (essentially) no chance of seeing what we observed if the company's claim is true, and so there is very strong evidence against the company's claim.

Chapter 8

Confidence Intervals

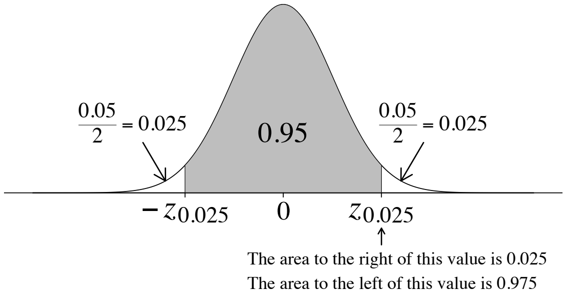
J.B.'s strongly suggested exercises: 1, 2, 4, 5, 7, 8, 9, 10, 11, 13, 14, 24, 26, 27, 29, 31, 32

8.1 Introduction

8.2 Interval Estimation of μ when σ is Known

1. (a) $z_{.025} = 1.96$ (see Figure 8.1), and the margin of error is $1.96 \times \frac{5}{\sqrt{22}} = 2.089$.
   (b) $z_{.05} = 1.645$, and the margin of error is $1.645 \times \frac{10}{\sqrt{10}} = 5.202$.
   (c) $z_{.005} = 2.576$, and the margin of error is $2.576 \times \frac{15}{\sqrt{18}} = 9.108$.
   (d) $\alpha = 1 - 0.632 = 0.368$. $\frac{\alpha}{2} = 0.184$. $z_{.184} = 0.90$, and the margin of error is $0.90 \times \frac{15}{\sqrt{18}} = 3.182$.
   (e) $\alpha = 1 - 0.212 = 0.788$. $\frac{\alpha}{2} = 0.394$. $z_{.394} = 0.27$, and the margin of error is $0.27 \times \frac{35}{\sqrt{50}} = 1.336$.

2. (a) 77.4% (since the area between -1.21 and 1.21 under the standard normal curve is 0.774 (see Figure 8.2).
   (b) 82.3% (since the area between -1.35 and 1.35 under the standard normal curve is 0.823).
   (c) 88.4% (since the area between -1.57 and 1.57 under the standard normal curve is 0.884).
   (d) 98.6% (since the area between -2.45 and 2.45 under the standard normal curve is 0.986).
   (e) 99.8% (since the area between -3.15 and 3.15 under the standard normal curve is 0.998).



(a) We need to find $z_{0.025}$.

```
> qnorm(.025)
[1] -1.959964
> qnorm(.975)
[1] 1.959964
```

(b) R tells us that $z_{0.025} = 1.959964$ (rounded to 6 decimal places).

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

(c) The standard normal table tells us that $z_{0.025} = 1.96$ (rounded to 2 decimal places).

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367

(d) The standard normal table tells us that $-z_{0.025} = -1.96$ (rounded to 2 decimal places).

Figure 8.1: Finding the appropriate z value for a 95% interval.

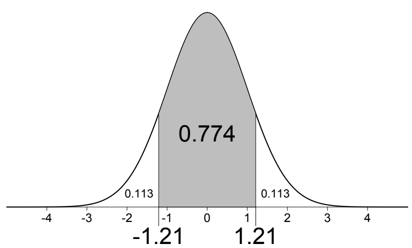


Figure 8.2: The confidence level is 0.774 (77.4%) if $z_{\alpha/2} = 1.21$.

3. (a) In statistical inference, we often construct confidence intervals for statistics. *False. We construct confidence intervals for parameters, and never for statistics.*
   (b) A confidence interval is a range of plausible values for a parameter. *True.*
   (c) The confidence level of an interval is determined by sample data. *False. We choose the confidence level.*
   (d) The point estimate of μ lies at the midpoint of the confidence interval for μ. *True. The point estimate of μ is the value of $\bar{X}$, which always falls at the midpoint of the interval.*
   (e) The confidence interval procedures of this section assume a normally distributed population, but this assumption becomes less important as the sample size increases. *True. (Because of the central limit theorem.)*

8.2.1 Interpretation of the Interval

4. (a) $25.7 \pm 1.96\frac{6.0}{\sqrt{200}}$, which works out to 25.7 ± 0.832 or (24.87, 26.53).
   (b) We can be 95% confident that the true mean BMI for female students at this university lies between 24.87 kg/m^2 and 26.53 kg/m^2.
   (c) $25.7 \pm 2.576\frac{6.0}{\sqrt{200}}$, which works out to 25.7 ± 1.093 or (24.61, 26.79).
   (d) We can be 99% confident that the true mean BMI for female students at this university lies between 24.61 kg/m^2 and 26.79 kg/m^2.

5. We can be 95% confident that the mean right ear length of all Caucasian Italian females between the ages of 18 and 30 lies between 60.91 mm and 62.89 mm. Or:

   We can be 95% confident that the true mean right ear length of Caucasian Italian females between the ages of 18 and 30 lies between 60.91 mm and 62.89 mm.

   (It can also be said that in repeated sampling, 95% of the 95% confidence intervals calculated in this manner would capture the true mean right ear length of Caucasian Italian females between the ages of 18 and 30.)

6. Statement (c) is a correct interpretation of the interval, and the rest of the statements are false.

8.2.2 What Factors Affect the Margin of Error?

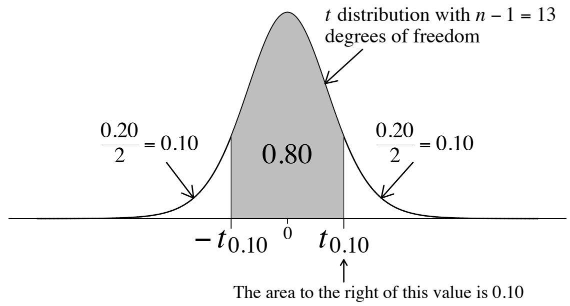
7. (a) An increase in the sample size decreases the margin of error.
   (b) An increases in the variance increases the margin of error.
   (c) An increase in $\bar{X}$ has no effect on the margin of error of the confidence interval.
   (d) An increase in the confidence level increases $z_{\alpha/2}$, which in turn increases the margin of error.

8.2.3 An Example

8.3 Confidence Intervals for μ When σ is Unknown

8.3.1 Introduction

8. (a) With 13 degrees of freedom, $t_{.10} = 1.350$ (see Figure 8.3). $SE(\bar{X}) = \frac{10}{\sqrt{14}} = 2.673$. $MOE = 1.350 \times 2.673 = 3.609$.



(a) We need to find $t_{0.10}$.

```
> qt(.1,13)
[1] -1.350171
> qt(.9,13)
[1] 1.350171
```

(b) R tells us that (with 13 DF) $t_{0.10}$ = 1.350171 (rounded to 6 decimal places).

	Area to the right								
df	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073

(c) The t table tells us that (with 13 DF) $t_{0.10} = 1.350$ (rounded to 3 decimal places).

Figure 8.3: Finding the appropriate t value for an 80% interval.

(b) With 21 degrees of freedom, $t_{.10} = 1.323$. $SE(\bar{X}) = \frac{10}{\sqrt{22}} = 2.132$. $MOE = 1.323 \times 2.132 = 2.821$.

(c) With 7 degrees of freedom, $t_{.05} = 1.895$. $SE(\bar{X}) = \frac{2500}{\sqrt{8}} = 883.8835$. $MOE = 1.895 \times 883.8835 = 1674.606$.

(d) With 7 degrees of freedom, $t_{.025} = 2.365$. $SE(\bar{X}) = \frac{2500}{\sqrt{8}} = 883.8835$. $MOE = 2.365 \times 883.8835 = 2090.031$.

(e) With 7 degrees of freedom, $t_{.005} = 3.499$. $SE(\bar{X}) = \frac{3.2}{\sqrt{8}} = 1.131$. $MOE = 3.499 \times 1.131 = 3.959$.

9. (a) The variance of the t distribution is greater than the variance of the standard normal distribution. *True.*

   (b) The mean of the t distribution is equal to the mean of the standard normal distribution. *True. (Both distributions have a mean of 0.)*

   (c) The t distribution has more area in the tails and a lower peak than the standard normal distribution. *True.*

   (d) As the degrees of freedom increase, the t distribution tends toward the standard normal distribution. *True.*

   (e) The t distribution is equivalent to the standard normal distribution if the degrees of freedom are at least 30. *False. Although the t distribution tends toward the standard normal distribution as the degrees of freedom increase,* *there are still meaningful differences at 30 degrees of freedom and above.*

10. The interval found using $z_{\alpha/2}$ would be narrower. Although narrower intervals are preferred, had we made this mistake we would be reporting incorrect results. (But if the sample size is large the mistake would have only a very small effect.)

8.3.2 Examples

11. (a) No, there is no indication of a violation of the normality assumption. The boxplot looks roughly symmetric, and there is no systematic curvature or outliers evident in the normal QQ plot.
    (b) The sample mean is the point estimate of the population mean: $\bar{X} = 50.91$ kgf. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{7.74}{\sqrt{32}} = 1.3683$.
    (c) Using software or a t table, we can find that with $n - 1 = 32 - 1 = 31$ degrees of freedom, $t_{0.025} = 2.0395$. The 95% interval is given by:

$$\begin{aligned} &\bar{X} \pm t_{0.025} SE(\bar{X}) \\ &50.91 \pm 2.0395 \times 1.3683 \\ &50.91 \pm 2.791 \end{aligned}$$

    which works out to (48.12, 53.70).
    (d) We can be 95% confident that the population mean grip strength of male students at this German university lies between 48.12 kgf and 53.70 kgf.

    The population is made up of all male students at this university at the time of the study. It wasn't clearly stated in the study how the volunteers were recruited, and the method of recruited could introduce bias. As an extreme example, if the volunteers were chosen from varsity sports teams, then that might severely bias the results (since athletes would likely tend to have a greater grip strength than non-athletes). We would need to know more information about the volunteer recruitment protocol to make an informed decision about whether this is a concern. In any event, the fact that the students volunteered for the study could possibly introduce bias, as volunteers might differ in a meaningful way from the general student population.

8.3.3 Assumptions of the One-Sample t Procedures

12. If the population is not normally distributed, then the *true* coverage probability of the confidence interval method can be very different from the nominal (stated) confidence level. In short, the reported results might be very misleading.

13. Strong skewness and outliers pose the biggest problems for the t procedure (and they are especially problematic if the sample size is small). If a distribution is roughly symmetric and unimodal (the distribution is "mound shaped"), then this type of violation of the normality assumption is not a big problem, even for small sample sizes. The effect of any violation of the normality assumption becomes less of a problem as the sample size increases.

8.4 Determining the Minimum Sample Size n

14. (a) $n \geq (\frac{1.645 \times 10}{5})^2 = 10.8$. We need a sample size of at least 11.
   (b) $n \geq (\frac{1.96 \times 10}{5})^2 = 15.4$. We need a sample size of at least 16. (Note that in minimum sample size calculations, we round *up* to the next largest whole number.)
   (c) $n \geq (\frac{1.96 \times 30}{5})^2 = 138.3$. We need a sample size of at least 139.
   (d) $n \geq (\frac{2.576 \times 10}{0.10})^2 = 66357.8$. We need a sample size of at least 66,358. (Note that the required sample size might be far larger than what is possible in practice. We may need to adjust our plans or abandon them if the calculation tells us that our wants are not practical.)
   (e) $n \geq (\frac{1.96 \times 4}{1})^2 = 61.47$. We need a sample size of at least 62.

8.5 Chapter Exercises

8.5.1 Basic Calculations

15. $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{16}} = 2.5$.

16. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{2.0}{\sqrt{16}} = 0.5$.

17. The population is assumed to be normally distributed, and the standard deviation is based on sample data, so to calculate the intervals we will use: $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$.

   (a) For the sample 3, 5, 12, $\bar{X} = 6.667$ and $s = 4.725816$. There are $n - 1 = 3 - 1 = 2$ degrees of freedom, and with 2 degrees of freedom, $t_{.025} = 4.303$. The 95% interval is given by:
   $6.667 \pm 4.303 \times \frac{4.7258}{\sqrt{3}}$, which works out to 6.667 ± 11.74.

(b) $9.0 \pm 4.303 \times \frac{13.5277}{\sqrt{3}}$, which works out to 9.0 ± 33.61.
(c) $2027.0 \pm 3.182 \times \frac{20.5589}{\sqrt{4}}$, which works out to 2027.0 ± 32.71.
(d) $-47.2 \pm 2.776 \times \frac{3.633180}{\sqrt{5}}$, which works out to -47.2 ± 4.51.

8.5.2 Concepts

18. D is the best interpretation. (The rest of the statements are false.)

19. (a) The number of intervals that contain μ will have a binomial distribution with $n = 100$ and $p = 0.95$.
    (b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.95^{100} = 0.994$.

20. (a) The central limit theorem tells us that the sampling distribution of $\bar{X}$ is approximately normal for large sample sizes, so in most cases when $n = 100$, methods based on the normal distribution are reasonable (even when there is some skewness). But there may be better methods of analysis, such as carrying out an inference procedure on a transformation of the data.
    (b) $\sigma_{\bar{X}} = \frac{1.0}{\sqrt{100}} = 0.10$.
    (c) We will use $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ to calculate the interval. Using a standard normal table or software, we can find the appropriate z value: $z_{0.04} = 1.75$. (The area to the right of 1.75 under the standard normal curve is 0.04, so using 1.75 in the interval formula will result in a 92% confidence level.) The interval is: $1.4 \pm 1.75 \times 0.10$, which works out to 1.4 ± 0.175.
    (d) The true value of μ is unknown. (We are constructing a confidence interval in order to estimate it.)
    (e) The parameter μ represents the *true* mean time to death for this type of rat after being infected with this virus (under the conditions of this experiment).

21. True. (This can be seen in the t table. For example, $t_{0.025} \to 1.96$ as the degrees of freedom increase.)

Sample Size (n)	Degrees of Freedom	$t_{.025}$
2	1	12.706
6	5	2.571
11	10	2.228
31	30	2.042
51	50	2.009
101	100	1.984
∞	∞	1.960

22. If we let $\bar{X}_1$ represent the mean of Sample 1 and $\bar{X}_2$ represent the mean of Sample

2, then we need to find $P(\bar{X}_1 - 1.96\frac{\sigma}{\sqrt{n}} < \bar{X}_2 < \bar{X}_1 + 1.96\frac{\sigma}{\sqrt{n}})$.

$$P(\bar{X}_1 - 1.96\frac{\sigma}{\sqrt{n}} < \bar{X}_2 < \bar{X}_1 + 1.96\frac{\sigma}{\sqrt{n}}) = P(-1.96\frac{\sigma}{\sqrt{n}} < \bar{X}_2 - \bar{X}_1 < 1.96\frac{\sigma}{\sqrt{n}})$$

Note that $\bar{X}_2 - \bar{X}_1$ has a mean of $\mu - \mu = 0$ and a standard deviation of $\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}} = \sqrt{2}\frac{\sigma}{\sqrt{n}}$. Standardizing:

$$\begin{aligned}
P(\bar{X}_1 - 1.96\frac{\sigma}{\sqrt{n}} < \bar{X}_2 < \bar{X}_1 + 1.96\frac{\sigma}{\sqrt{n}}) &= P(\frac{-1.96\frac{\sigma}{\sqrt{n}} - 0}{\sqrt{2}\frac{\sigma}{\sqrt{n}}} < Z < \frac{1.96\frac{\sigma}{\sqrt{n}} - 0}{\sqrt{2}\frac{\sigma}{\sqrt{n}}}) \\
&= P(-\frac{1.96}{\sqrt{2}} < Z < \frac{1.96}{\sqrt{2}}) \\
&= 0.834
\end{aligned}$$

The probability that the sample mean of one sample falls within the 95% confidence interval for μ found from another sample of the same size is only 0.834.

23. We are given that $\hat{\theta} = 2.00$ and $\sigma_{\hat{\theta}} = 3.00$ (the standard deviation of the sampling distribution of $\hat{\theta}$ is 3.00). Since the estimator $\hat{\theta}$ is normally distributed, and its standard deviation is known, the margin of error is $z_{\alpha/2} \times \sigma_{\hat{\theta}} = 1.96 \times 3 = 5.88$, and the 95% interval for θ is 2.00 ± 5.88.

24. Test your conceptual understanding: Which of the following statements are true? You should be able to explain *why* a statement is true or *why* a statement is false.

    (a) The standard error of a statistic decreases as the sample size decreases. *False. The standard error increases as the sample size decreases.*
    (b) The one-sample t procedures are robust to violations of the normality assumption. *True. The t procedures work quite well in a variety of non-normal situations.*
    (c) The t procedures do not perform well when there is strong skewness or outliers in the data, especially for small sample sizes. *True.*
    (d) The width of a confidence interval for μ decreases as the sample size decreases. *False. The width of a confidence interval increases as the sample size decreases.*
    (e) The width of a confidence interval for μ decreases as the variance increases. *False. The width of a confidence interval increases as the variance increases.*

25. Test your conceptual understanding: Which of the following statements are true? You should be able to explain *why* a statement is true or *why* a statement is false.

(a) In practical problems, σ is usually known and thus we usually use z procedures instead of t procedures. *False. The vast majority of the time σ is an unknown quantity.*

(b) Standard errors are the estimated standard deviations of parameters in repeated sampling. *False. A standard error is the estimated standard deviation of a statistic in repeated sampling.*

(c) If the assumptions of a procedure used to calculate a 90% confidence interval are violated, the true confidence level of the interval may be very different from 90%. *True. If the assumptions are false the true confidence level may be very different from what is stated.*

(d) All else being equal, the width of a confidence interval decreases as the confidence level increases. *False. If we want greater confidence that the interval captures the parameter, we must cover more ground and have a wider interval.*

(e) All else being equal, the margin of error decreases as the confidence level increases. *False. The margin of error increases as the confidence level increases.*

(f) All else being equal, the standard error of the sample mean decreases as the confidence level of the interval increases. *False.* $SE(\bar{X}) = \frac{s}{\sqrt{n}}$, *which does not involve the confidence level.*

(g) All else being equal, the confidence level of an interval increases as the sample size increases. *False. The confidence level* $(1-\alpha)$ *and sample size* (n) *are not related.*

8.5.3 Applications

26. (a) Both the boxplot and the normal QQ plot show indicate right-skewness (which is typical of BMI data). This skewness is not a serious problem in this scenario, as the skewness is not very strong, and the sample size is fairly large (94). We can use the t procedures. (We will use inference procedures on the mean in this question, which is perfectly valid. But because of the skewness, in real world scenarios the *median* BMI is often reported instead of the *mean* BMI.)

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 23.3$. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{5.4}{\sqrt{94}} = 0.5570$.

(c) Using software, we can find that with 93 degrees of freedom, $t_{0.025} = 1.9858$. (If we must use a table, we may have to look up 90 df and find the t value of 1.987.)

$$\bar{X} \pm t_{0.025} SE(\bar{X})$$
$$23.3 \pm 1.9858 \times 0.5570$$
$$23.3 \pm 1.106$$

which works out to (22.2, 24.4). (Had we used the t value of 1.987 from the table, we would have gotten the same answer to 1 decimal place.)

(d) The interpretation of the interval is a little tricky here, due to the complexities of the sampling design. The sample is not truly a simple random sample from any population, so one could argue that we don't really have a confidence interval for anything. But sampling designs like this are very common in practice (when human consent is required, there are always going to be biases present). So I'll give a not-quite-true-but-somewhat-reasonable interpretation, then discuss the biases:

We can be 95% confident that μ, the true mean BMI for 14-year old boys in this area of rural Texas, lies between 22.2 and 24.4 kg/m^2.

The sample is drawn from a school district in north central Texas, so, loosely, the population is all 14 year-old boys in this area of north central Texas. However, there are some sampling biases that make this interpretation not quite true:

- The sample was drawn from only 6 schools in that area. Students in these schools may differ in a meaningful way from students at other schools in the area (and from children that are not enrolled at a school).
- Parents needed to provide informed consent in order for students to participate. Students whose parents provide informed consent may differ in a meaningful way from those whose parents do not.
- Students needed to provide self-assent (the student had to agree to be part of the study). Students who agree to participate may differ in a meaningful way from those who do not.
- Students who participated were given a t-shirt. This could introduce bias, as it is conceivable that there is a relationship between desire for the t-shirt and BMI.

27. (a) No, there is no indication of a violation of the normality assumption. The boxplot looks symmetric, and there is no systematic curvature or outliers evident in the normal QQ plot. A minor violation of the normality assumption would not be a big deal in this scenario, as the sample size is not small ($n = 51$). We can use the t procedures.

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 1.65$ mm/s. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.62}{\sqrt{51}} = 0.0868$.

(c) Using software or a t table, we can find that with $n - 1 = 51 - 1 = 50$ degrees of freedom, $t_{0.025} = 2.009$. (If we must use a table, we may have to look up 90 df and find a t value of 1.987.)

$$\begin{gathered}\bar{X} \pm t_{0.025} SE(\bar{X}) \\ 1.65 \pm 2.009 \times 0.0868 \\ 1.65 \pm 0.174\end{gathered}$$

which works out to (1.48, 1.82).

(d) We can be 95% confident that the population mean speed of tandem running lies between 1.48 and 1.82 mm/s.
It wasn't clearly stated how the tandem running pairs were chosen, but it is reasonable to think that these results apply to *Temnothorax albipennis* ants (with a young experienced ant leading a young inexperienced ant) under the conditions of the study. The true mean may be different if the conditions change, or if a different type of ant is studied.

28. (a) We need a simple random sample from a normally distributed population. The skeletons in the Athens Collection do not truly represent a simple random from any population, and to come up with an informed opinion we would need to look in greater detail at how this sample was gathered. However, in a loose way, we can think of this sample of skeletons as a random sample of skeletons from men who lived near Athens near the end of the 20th century.

As far as normality is concerned, we don't have the actual data to investigate that, but it is likely that the diameters are approximately normally distributed. The sample size ($n = 24$) is not very large, so normality is a concern. If we had access to the data, we should investigate the normality assumption with a normal QQ plot.

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 10.38$ mm/s. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.63}{\sqrt{24}} = 0.1286$.

(c) Using software or a t table, we can find that with $n - 1 = 24 - 1 = 23$ degrees of freedom, $t_{0.025} = 2.069$.

$$\begin{gathered}\bar{X} \pm t_{0.025} SE(\bar{X}) \\ 10.38 \pm 2.069 \times 0.1286 \\ 10.38 \pm 0.266\end{gathered}$$

which works out to (10.11, 10.65).

(d) We can be 95% confident that the true mean crown diameter of upper M1 molars lies between 10.11 and 10.65 mm. The population to which this conclusion applies depends on how this sample of skeletons was drawn. In a loose way, we can think of the population as adult males who lived near Athens near the end of the 20th century, but we would need to investigate the sampling design in greater detail to reach an informed opinion.

29. (a) We would need the sample to be a simple random sample from a normally distributed population. The sample of female blowflies is not truly a simple random sample of blowflies from the area (they were captured in traps, and there could conceivably be factors influencing the types of fly that get trapped). We also need the number of larvae in flies of this type to be normally distributed. (But since the sample size is not small ($n = 49$), the normality assumption is not very important.)

(b) The largest observation is roughly 1.5 standard deviations greater than the mean, and the smallest is roughly 2 standard deviations below the mean. So there are not any extreme values (outliers) that would affect the t procedures. This information makes the use of the t procedures more reasonable.

(c) The sample mean is the point estimate of the population mean: $\bar{X} = 33.4$ mm/s. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{7.0}{\sqrt{49}} = 1.0$.

(d) Using software, we can find that with 48 degrees of freedom, $t_{0.025} = 2.0106$. (If we must use a table, we may have to look up 45 df and find the t value of 2.014.)

$$\begin{aligned} &\bar{X} \pm t_{0.025} SE(\bar{X}) \\ &33.4 \pm 2.0106 \times 1.0 \\ &33.4 \pm 2.01 \end{aligned}$$

which works out to (31.4, 35.4). (Had we used the t value of 2.014 from the table, we would have gotten the same answer to 1 decimal place.)

(e) We can be 95% confident that the true mean number of larvae carried by female *C. varifrons* blowflies in this field lies between 31.4 and 35.4.

There are possible biases introduced by the sampling mechanism. Flies with certain characteristics (old or large flies, for example) may be more or less likely to be captured. These characteristics may also be related to the number of larvae that they carry. (For example, it is conceivable that older flies would be less likely to be captured. If older flies also tend to carry fewer larvae, then the interval would tend to be too large (since older flies would be underrepresented in the sample).)

Also, the interval only applies to this type of fly under the conditions of the study. If we were to use this interval to estimate the mean in other situations (different locations, temperatures, humidity, or for a different type of fly), then the interval may be very far off and very misleading.

30. (a) We would need the sample of flies to behave like a simple random sample of *C. varifrons* blowflies. We would also need the rate of deposition to be approximately normally distributed. We would need to investigate the study in greater detail to form an opinion about the sample of flies that they used.

Since the sample size is not small ($n = 42$), it would be reasonable to use the t procedures, unless the data showed strong skewness or outliers.

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 0.46$ mm/s. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.32}{\sqrt{42}} = 0.049$.

(c) Using software, we can find that with 41 degrees of freedom, $t_{0.025} = 2.0195$. (If we must use a table, we may have to look up 40 df and find the t value of 2.021.)

$$\begin{aligned}&\bar{X} \pm t_{0.025} SE(\bar{X})\\&0.46 \pm 2.0195 \times 0.049\\&0.46 \pm 0.100\end{aligned}$$

which works out to (0.36, 0.56). (Had we used the t value of 2.021 from the table, we would have gotten the same answer to 2 decimal places.)

(d) We can be 95% confident that the population mean rate of larviposition lies between 0.36 and 0.46 larvae per second.

This interval applies to this type of blowfly under the conditions of this experiment. If the conditions of the experiment were to change, then the true mean rate of larviposition may change. Biases may have been introduced by the choice of *C. varifrons* blowfly used in this experiment. (It is not clear how the *C. varifrons* blowflies used in the experiment were chosen.)

31. (a) To draw conclusions about all hash browns made in U.S. locations of this fast-food chain, we need the hash browns in the sample to be a random sample from the population of hash browns in these U.S. locations. Little is said in the U.S. Nutrient Database about how the sample was collected, but the people that compile this database are professionals, and it is unlikely that this sample has obvious biases. (Discussion of possible bias will be discussed below in the answer to Question 31d.)

For the t procedure to be reasonable, we need the amount of sodium in hash browns of this type to be normally distributed. This is especially true here, since the sample size is so small ($n = 6$). It is very likely that the amount of sodium in hash browns of this type is approximately normally distributed, but we cannot tell from the given information.

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 324.8$mg. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{40.0}{\sqrt{6}} = 16.330$.

(c) Using software or a t table, we can find that with 5 degrees of freedom, $t_{0.025} = 2.571$.

$$\begin{aligned}&\bar{X} \pm t_{0.025} SE(\bar{X})\\&324.8 \pm 2.571 \times 16.330\\&324.8 \pm 41.98\end{aligned}$$

which works out to (282.82, 366.78).

(d) Assuming normality, and assuming that the sample of hash browns can be thought of as a random sample of hash browns from U.S. locations of this chain:
We can be 95% confident that the true mean sodium content in servings of hash browns from this fast-food chain lies between 282.82 mg and 366.78 mg.

There are many sources of possible bias, depending on how the sample was drawn. The distribution of sodium may differ from location to location, as preparation methods might differ slightly between locations. (The fast-food chain in question is a very large chain with common preparation methods, so this is unlikely to be major influence, but there still could be an effect.) If the sample was drawn from an individual location, and we were trying to draw inferences about all U.S. restaurants, this could easily result in bias. If the sample of 6 hash browns was drawn from a one batch of hash browns, then it would be very unwise to try to draw conclusions about all hash browns from this type of sample.

(e) Yes, the company's claimed value of 310 mg is a plausible value of the population mean, since it falls within the 95% confidence interval.

32. (a) To draw conclusions about black women in the Manchester area, we need the women in the sample to be a random sample from the population of black women in the Manchester area. The sample was drawn from women attending a sexual health clinic, and these women might differ in a meaningful way from the general population of black women in the area. (Sources of possible bias will be discussed below in the answer to Question 32d.)

For the t procedure to be reasonable, we also need the 2D:4D ratios to be normally distributed. The article stated that the distribution was close to normal, so the normality assumption is reasonable. The sample size is 46, which is not small, so the t procedures would still be reasonable even if there were a little skewness or other issues with mild non-normality.

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 0.963$. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.034}{\sqrt{46}} = 0.0050$.

(c) Using software or a t table, we can find that with 45 degrees of freedom, $t_{0.025} = 2.014$.

$$\bar{X} \pm t_{0.025} SE(\bar{X})$$
$$0.963 \pm 2.014 \times 0.0050$$
$$0.963 \pm 0.010$$

which works out to (0.953, 0.973).

(d) We can be 95% confident that the true mean 2D:4D ratio (of the left hand) of black women attending this sexual health clinic in Manchester lies between 0.953 and 0.973.

There are many sources of possible bias, depending on the conclusion one wishes to make. Some things to consider:

- The sample of women was drawn from black women attending a sexual health clinic in Manchester, UK. It is reasonable to think that the conclusions apply to black women who attend this clinic.
- It is likely reasonable to think that these results apply to all black women in the Manchester, UK area, but we need to be cautious with a conclusion like this, since black women attending this sexual health clinic might differ in a meaningful way from the general population of black women in the area.
- Although these results give some indication of the 2D:4D ratio for black women in general, generalizing beyond the area from which the sample was drawn is always dubious and may lead to misleading conclusions.
- As is often the case involving studies with human participants, individuals were included in this study only if they provided written consent. Women who consent to studies like this might differ in a meaningful way from those who do not.

33. (a) We would need these intimate partner homicides to be a simple random sample from a (theoretical) population of intimate partner homicides. We also need the ages of the perpetrators of intimate partner homicide to be normally distributed. These assumptions will not be perfectly justified in this case. There are some sources of possible bias (discussed below in the answer to Question 33d). The ages will not be perfectly normally distributed, but the sample size is not too small ($n = 36$), so unless there is strong skewness or outliers, the t procedure should work reasonably well.

(b) The sample mean is the point estimate of the population mean: $\bar{X} = 43.8$ years. $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{15.5}{\sqrt{36}} = 2.583$.

(c) Using software or a t table, we can find that with 41 degrees of freedom, $t_{0.025} = 2.030$.

$$\begin{aligned} &\bar{X} \pm t_{0.025} SE(\bar{X}) \\ &43.8 \pm 2.030 \times 2.583 \\ &43.8 \pm 5.24 \end{aligned}$$

which works out to (38.6, 49.0).

(d) We can be 95% confident that the true mean age of perpetrators of intimate partner homicides in Denmark during the time period 1983–2007 lies between 38.6 to 49.0 years.

One source of bias is that the sample of homicides included only those scenarios that were determined to be intimate partner homicides. (In reality, not all homicides are correctly labelled as such, and not all intimate partner homicide perpetrators are properly labelled as such.) It is conceivable that the age of perpetrators in which the homicide is not classified properly tends to be older or younger than the age of perpetrators in homicides classified as intimate partner homicides. So the interval applies only to the population of scenarios in which the death was labelled as an intimate partner homicide. This population would differ (in a small but meaningful way) from the population of scenarios that were truly intimate partner homicides.

8.5.4 Extra Practice Questions

34. (a) There are no obvious problems here. The points fall close to a straight line, with no obvious curvature or other problems (indicating that the data is approximately normally distributed). It is reasonable to assume normality and use the t procedure.
    (b) The standard deviation is based on sample data, so the interval is found using the formula $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$. With $15 - 1 = 14$ degrees of freedom, $t_{.025} = 2.145$. The endpoints of the interval are: $295.6 \pm 2.145 \times \frac{10.2}{\sqrt{15}}$, which works out to 295.6 ± 5.65 or $(289.95, 301.25)$.
    (c) The second statement (ii) is correct and the rest of the statements are false.

35. Of the given options, D is the best one. (The rest of the statements are false.)

36. We can be 95% confident that the true mean wait time for all calls to this call centre lies between 2.88 and 3.12 minutes.

37. (a) Plotting a normal quantile-quantile plot can give a good indication of whether or not the data is approximately normal.
    (b) The interval is $6.8 \pm 2.365 \times \frac{0.32}{\sqrt{8}}$, which works out to 6.8 ± 0.2676, or $(6.53, 7.07)$.
    (c) We can be 95% confident that the true mean fuel consumption of this type of car under these conditions lies between 6.53 and 7.07 l/100k.

38. (a) The value of $\bar{X}$ is the point estimate of μ. Here, $\bar{X} = 29.3$.
    (b) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{11.9}{\sqrt{26}} = 2.3338$.
    (c) $29.3 \pm 2.060 \times 2.3338$, which works out to 29.3 ± 4.81 or $(24.49, 34.11)$.
    (d) $1.708 \times \frac{11.9}{\sqrt{26}} = 3.986$.
    (e) 29.3 ± 3.986 or $(25.31, 33.29)$.
    (f) We can be 90% confident that the true mean number of facial tissues used per day by this type of cold sufferer under these conditions lies between 25.31 and

33.29.

39. We can be 95% confident that the true mean melting temperature of this new type of oil lies between 34.34 and 34.71 degrees Celsius.

40. (a) $3.2 \pm 2.262 \times \frac{0.6}{\sqrt{10}}$, which works out to 3.2 ± 0.429 or (2.77, 3.63).
    (b) We can be 95% confident that the true mean cadmium concentration in mushrooms of this type in this area lies between 2.77 and 3.63 mg/kg of dry weight.
    (c) We should never simply assume the data is normally distributed; this assumption should be investigated. We can investigate the normality assumption with a normal quantile-quantile plot. If the cadmium concentrations are not normally distributed, then what we are calling a 95% confidence interval may have a true coverage probability that is far different from 95%. (Our stated confidence level may be very different from the true confidence level.)

41. (a) $9.2 \pm 2.110 \times \frac{0.38}{\sqrt{18}}$, which works out to 9.2 ± 0.189 or (9.01, 9.39).
    (b) The first statement is true and the rest are false.

42. (a) 61.1. (The value of $\bar{X}$ is a point estimate of μ.)
    (b) $\frac{61.8-60.4}{2} = 0.7$. (This is half the width of the interval.)
    (c) The 95% margin of error was found to be 0.7. Since the 95% margin of error is $1.96\sigma_{\bar{X}}$, this implies $\sigma_{\bar{X}} = \frac{0.7}{1.96} = 0.3571$. To calculate the 93% interval, we first need to find the appropriate $z_{\alpha/2}$ value: $z_{0.035} = 1.81$ (the area to the right of 1.81 under the standard normal curve is 0.035). The 93% interval is given by $61.1 \pm 1.81 \times 0.3571$, which works out to 61.1 ± 0.646.

Chapter 9

Hypothesis Testing

J.B.'s strongly suggested exercises: 1, 5, 6, 10, 12, 13, 15, 19, 21, 33, 37, 38, 41, 48, 59, 61, 62, 63, 64, 65, 66

9.1 Introduction

1. (a) $H_0 : \mu_M = \mu_F$ (the true mean resting heart rate is the same for male and female Olympic athletes).
   $H_a : \mu_M \neq \mu_F$ (the true mean resting heart rate is different between male and female Olympic athlet
   (b) $H_0 : p_1 = p_2 = p_3 = p_4$ (the six month survival rate is the same for all four treatments).
   $H_a :\ p_i \neq p_j$ for at least one i, j combination (the six month survival rates are not all equal).
   (c) $H_0 : \mu = 4.0$ (the true mean saturated fat content is 4.0 grams).
   $H_a : \mu > 4.0$ (the true mean saturated fat content is greater than 4.0 grams).

   It would also be acceptable to express the hypotheses this way:
   $H_0 : \mu \leq 4.0$ (the true mean saturated fat content is no more than 4.0 grams).
   $H_a : \mu > 4.0$ (the true mean saturated fat content is greater than 4.0 grams).
   But when carrying out the test we will need to test a specific value, so in this text we will always express the null hypothesis as an equality.
   (d) $H_0 : p = 0.98$ (the true proportion of satisfied customers is 0.98).
   $H_a : p < 0.98$ (the true proportion of satisfied customers is less than 0.98).
   (e) $H_0 :\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ (the true variances are all equal).
   $H_a : \sigma_i^2 \neq \sigma_j^2$ for at least one i, j combination (the true variances are not all equal).

9.2 The Logic of Hypothesis Testing

9.3 Hypothesis Tests for μ when σ is Known

9.3.1 Constructing Appropriate Hypotheses

2. We would *never* test H_0: $\bar{X} = 2$. (It is not an appropriate hypothesis, since hypotheses never involve statistics.) We test hypotheses about *parameters* (μ, for example).

9.3.2 The Test Statistic

3. In this question we are sampling from a normally distributed population where the population standard deviation σ is known, so we should use the Z statistic: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$.

   (a) $Z = \frac{15.5-20}{5/\sqrt{10}} = -2.85$.
   (b) $Z = \frac{15.5-20}{5/\sqrt{40}} = -5.69$.
   (c) $Z = \frac{35.5-30}{5/\sqrt{20}} = 4.91$.
   (d) $Z = \frac{65.5-30}{5/\sqrt{20}} = 31.75$.

4. $\frac{\bar{X}-12}{6/\sqrt{75}}$ is the Z statistic: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$. When we are sampling from a normally distributed population and the null hypothesis is true, the Z statistic has the standard normal distribution.

9.3.3 The Rejection Region Approach to Hypothesis Testing

5. (a) Reject H_0 if $|Z| \geq 1.645$.
   (b) Reject H_0 if $Z \leq -1.28$.
   (c) Reject H_0 if $Z \geq 1.28$.
   (d) Reject H_0 if $|Z| \geq 2.88$.
   (e) Reject H_0 if $Z \leq -2.65$.
   (f) Reject H_0 if $Z \geq 2.65$.

6. (a) The significance level of a test is the probability of rejecting the null hypothesis, given the null hypothesis is true. *True.*

(b) If we reject the null hypothesis, then we know that the null hypothesis is false. *False. There is an important difference between the result of the test and the (unknown) underlying reality. Whatever the result of the hypothesis test, we still do not know the underlying reality.*

(c) If we reject the null hypothesis, then we know that the alternative hypothesis is false. *False.*

(d) If we do not reject the null hypothesis, then we know that the null hypothesis is true. *False.*

(e) If we do not reject the null hypothesis, then we know that the alternative hypothesis is true. *False.*

9.3.4 P-values

7. (a) p-value $= 0.0058$ (twice the area to the right of 2.76 under the standard normal curve).
   (b) p-value $= 0.9971$ (the area to the left of 2.76 under the standard normal curve).
   (c) p-value $= 0.0029$ (the area to the right of 2.76 under the standard normal curve).
   (d) p-value $= 0.0818$ (twice the area to the left of -1.74 under the standard normal curve).
   (e) p-value $= 0.0409$ (the area to the left of -1.74 under the standard normal curve).
   (f) p-value $= 0.9591$ (the area to the right of -1.74 under the standard normal curve).

8. (a) Reject H_0, since p-value $< \alpha$.
   (b) Reject H_0, since p-value $< \alpha$.
   (c) Do not reject H_0, since p-value $> \alpha$.
   (d) Do not reject H_0, since p-value $> \alpha$.
   (e) Reject H_0, since p-value $< \alpha$.

9. A is correct, and the rest of the statements are false. The p-value is the probability, given the null hypothesis is true, of obtaining a test statistic with at least as much evidence against the null hypothesis as what was observed in the sample.

10. (a) A p-value is a probability. *True.*
    (b) We reject the null hypothesis when the p-value is less than or equal to the significance level. *True.*
    (c) The p-value of a test is the probability, given H_0 is true, of obtaining the observed value of the test statistic or a value with even greater evidence against H_0 and in favour of H_a. *True.*

(d) The p-value of a test depends on sample data and on the null and alternative hypotheses. *True.*

(e) The p-value of a two-sided test will be greater than 1 if the null hypothesis is true. *False. A p-value is a probability and as such cannot be greater than 1.*

11. (a) The p-value of a test will be less than 0.05 if the null hypothesis is false. *False. The p-value can be any value between 0 and 1, whether or not the null hypothesis is true. Keep in mind that there is a fundamental difference between the conclusion from a test and the (unknown) underlying reality.*

(b) The p-value of a test is the probability that the null hypothesis is true. *False.*

(c) If the p-value of a hypothesis test with a two-sided alternative is equal to exactly 1, then we can be certain the null hypothesis is true. *False.*

(d) If the p-value of a hypothesis test with a two-sided alternative is equal to exactly 1, then we can be certain the null hypothesis is false. *False.*

(e) If the p-value of a test is less than 0 that means there is extremely strong evidence against the null hypothesis. *False. A p-value is a probability and as such it cannot be less than 0. If we find a negative p-value, that means we made a mistake.*

9.4 Examples

12. (a) H_0: $\mu = 3.0$ mg/kg dw. (The true mean cadmium concentration in caps of *Boletus edulis* in this area is 3.0 mg/kg of dry weight. (The true mean cadmium concentration is at the tolerable value.).)

H_a: $\mu > 3.0$ mg/kg dw. (The true mean cadmium concentration in caps of *Boletus edulis* in this area is greater than 3.0 mg/kg of dry weight. (The true mean cadmium concentration exceeds the tolerable value.).)

It would also be acceptable to express the hypotheses this way:
H_0: $\mu \leq 3.0$. (The true mean cadmium concentration in caps of *Boletus edulis* in this area is no more than 3.0 mg/kg of dry weight. (The true mean cadmium concentration is no more than the tolerable value.).)
$H_a : \mu > 3.0$ (The true mean cadmium concentration in caps of *Boletus edulis* in this area is greater than 3.0 mg/kg of dry weight. (The true mean cadmium concentration exceeds the tolerable value.).)
In some ways this set of hypotheses makes more sense, and is more fitting with the spirit of the test. But when carrying out the test we will need to test a specific value, so in this text we will always express the null hypothesis as an equality.

(b) $Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{9.26-3}{5.00/\sqrt{29}} = 6.74$.

(c) Since the alternative hypothesis is H_a: $\mu > 3.0$, the p-value is the area to the right of 6.74 under the standard normal curve. This area is very, very small (see Figure 9.1).

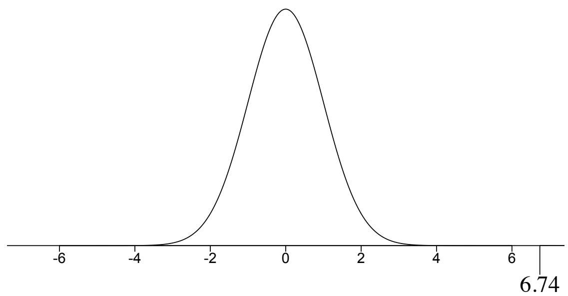


Figure 9.1: The p-value is the area to the right of 6.74 under the standard normal curve.

Using software: p-value $= 7.8 \times 10^{-8}$. Using a standard normal table, we could say only that the p-value is very close to 0 (or we may give a range, such as p-value < 0.0000003).

(d) The p-value is less than the given significance level of $\alpha = 0.05$, so we can reject the null hypothesis at $\alpha = 0.05$.

(e) There is statistically significant evidence (at $\alpha = 0.05$) against the null hypothesis that $\mu = 3.0$ mg/kg dw. So there is very strong evidence that the population mean μ is greater than 3.0 mg/kg dw. This means that there is statistically significant evidence that the true mean concentration of cadmium in caps of *Boletus edulis* mushrooms in this area is greater than the Slovenian tolerable value of 3.0 mg/kg dw.

13. (a) H_0: $\mu = 3.35$ kg. (The true mean weight of full-term Hispanic newborns in the U.S. in 2013 is 3.35 kg – the same mean as for full-term newborns of the general population.)
H_a: $\mu \neq 3.35$ kg. (The true mean weight of full-term Hispanic newborns in the U.S. in 2013 differs from 3.35 kg. That is, the true mean weight differs from that of full-term newborns of the general population.)

(b) $Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{3.318-3.35}{0.47/\sqrt{100}} = -0.681$.

(c) 0.496. The alternative hypothesis is two-sided, so the p-value is double the area in the tail, beyond the observed value of the test statistic. In this case, the p-value is double the area to the left of -0.681 under the standard normal curve (see Figure 9.2).
The area to the left of -0.681 under the standard normal curve is 0.248, and so the p-value is 0.496.

(d) The p-value is greater than the given significance level of $\alpha = 0.05$, so we do not reject the null hypothesis at this level of significance. (The p-value is in fact *much* larger than 0.05, so we would not reject the null hypothesis at

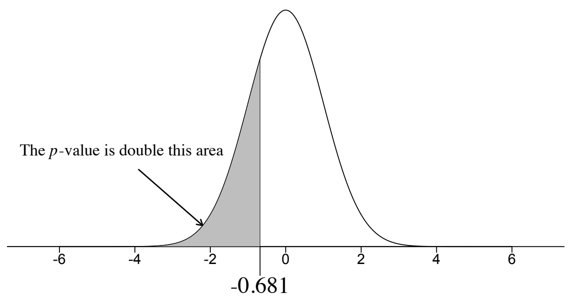


Figure 9.2: The p-value is double the area to the left of -0.681 under the standard normal curve.

any reasonable significance level, but we used $\alpha = 0.05$ since it was the given significance level for this problem.)

(e) Based on this sample, there is no evidence that the true mean weight of full-term Hispanic newborns in 2013 differs from 3.35 kg (the mean weight of all full-term U.S. newborns that year). We can say that the data is consistent with full-term Hispanic newborns having the same mean weight as that of full-term newborns in the general population. (N.B. We cannot say that they *are* equal, just that we do not have any evidence of a difference.)

9.5 Interpreting the p-value

9.5.1 The Distribution of the p-value When H_0 is True

14. (a) The p-value will have a uniform distribution on the interval (0,1).
    (b) If the null hypothesis is true, on average the p-value will equal 0.5.

9.5.2 The Distribution of the p-value When H_0 is False

15. (a) A p-value of 2.9×10^{-48} gives *extremely* strong evidence against the null hypothesis. There is almost no chance of observing what was observed if the null hypothesis were true.
    (b) A p-value of 0.0008 gives very strong evidence against the null hypothesis. There is less than a one-in-a-thousand chance of observing what was observed if the null hypothesis were true.
    (c) A p-value of 0.02 gives strong, but not overwhelming, evidence against the null hypothesis.

(d) A p-value of 0.07 gives a little evidence against the null hypothesis, but it is not strong (and not significant at the commonly used α level of 0.05).
(e) A p-value of 0.12 gives possibly a hint of evidence against the null hypothesis, but nothing that is significant at any of the usual significance levels.
(f) A p-value of 0.45 gives no evidence against the null hypothesis.
(g) A p-value of 0.88 gives no evidence against the null hypothesis.

9.6 Type I Errors, Type II Errors, and the Power of a Test

16. (a) Here the null hypothesis is false and we will reject it, so we will make neither error. (Our decision is consistent with the underlying reality.)
(b) Here the null hypothesis is false and we will not reject it, so we will make a Type II error.
(c) Here the null hypothesis is true and we will reject it, so we will make a Type I error.
(d) Here the null hypothesis is true and we will not reject it, so we will make neither error. (Our decision is consistent with the underlying reality.)

17. It all boils down to the balance between Type I and Type II errors. The significance level α is the probability of a Type I error (given the null hypothesis is true). If a Type I error has major repercussions (thousands of people will die, for example), but a Type II error is not as costly, then we should choose a very small value for α. If a Type I error is not very costly relative to a Type II error, then a larger choice of α is reasonable.

9.6.1 Calculating Power and the Probability of a Type II Error

18. (a) Reject H_0 if $\bar{X} \leq 26.710$. With the alternative hypothesis of H_a: $\mu < 30$ and $\alpha = 0.05$, we reject the null hypothesis if $Z \leq -1.645$. Expressing the rejection region in terms of $\bar{X}$:

$$\begin{aligned} Z &\leq -1.645 \\ \frac{\bar{X} - 30}{20/\sqrt{100}} &\leq -1.645 \\ \bar{X} &\leq 30 + \frac{20}{\sqrt{100}}(-1.645) \\ \bar{X} &\leq 26.710 \end{aligned}$$

(b) 0.259. The power of the test is the probability of rejecting the null hypothesis, given $\mu = 28$:

$$\begin{aligned}\text{Power} = P(\bar{X} \leq 26.710|\mu = 28) &= P(Z \leq \frac{26.710 - 28}{20/\sqrt{100}})\\ &= P(Z \leq -0.645)\\ &= 0.259\end{aligned}$$

(c) 0.639. The power of the test is the probability of rejecting the null hypothesis, given $\mu = 26$:

$$\begin{aligned}\text{Power} = P(\bar{X} \leq 26.710|\mu = 26) &= P(Z \leq \frac{26.710 - 26}{20/\sqrt{100}})\\ &= P(Z \leq 0.355)\\ &= 0.639\end{aligned}$$

19. (a) Reject H_0 if $\bar{X} \geq 818.61$. With the alternative hypothesis of H_a: $\mu > 800$ and $\alpha = 0.01$, we reject the null hypothesis if $Z \geq 2.326$.. (If you use a table to find the z value, you will only be able to find it to 2 decimal places: $Z \geq 2.33$. There will be a little rounding error.) Expressing the rejection region in terms of $\bar{X}$:

$$\begin{aligned}Z &\geq 2.326\\ \frac{\bar{X} - 800}{160/\sqrt{400}} &\geq 2.326\\ \bar{X} &\geq 800 + \frac{160}{\sqrt{400}}(2.326)\\ \bar{X} &\geq 818.61\end{aligned}$$

(b) 0.140. The power of the test is the probability of rejecting the null hypothesis, given $\mu = 810$:

$$\begin{aligned}\text{Power} = P(\bar{X} \geq 818.61|\mu = 810) &= P(Z \geq \frac{818.61 - 810}{160/\sqrt{400}})\\ &= P(Z \geq 1.076)\\ &= 0.140\end{aligned}$$

(c) 0.996. The power of the test is the probability of rejecting the null hypothesis, given $\mu = 840$:

$$\begin{aligned}\text{Power} = P(\bar{X} \geq 818.61|\mu = 840) &= P(Z \geq \frac{818.61 - 840}{160/\sqrt{400}})\\ &= P(Z \geq -2.674)\\ &= 0.996\end{aligned}$$

20. (a) Reject H_0 if $\bar{X} \leq 11.674$ or $\bar{X} \geq 18.326$.
Recall that for a two-sided alternative hypothesis and $\alpha = 0.05$, we reject H_0 if $Z \leq -1.96$ or $Z \geq 1.96$. We reject H_0 if:

$$Z \leq -1.96 \qquad \text{or} \qquad Z \geq 1.96$$

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -1.96 \qquad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq 1.96$$

$$\frac{\bar{X} - 15}{12/\sqrt{50}} \leq -1.96 \qquad \frac{\bar{X} - 15}{12/\sqrt{50}} \geq 1.96$$

$$\bar{X} \leq 11.674 \qquad \bar{X} \geq 18.326$$

(b) The power of the test if $\mu = 16$ is the total area in these two regions under the sampling distribution of $\bar{X}$.

$$\begin{aligned}
\text{Power} &= P(\bar{X} \leq 11.674|\mu = 16) + P(\bar{X} \geq 18.326|\mu = 16) \\
&= P(Z \leq \frac{11.674 - 16}{12/\sqrt{50}}) + P(Z \geq \frac{18.326 - 16}{12/\sqrt{50}}) \\
&= P(Z \leq -2.549) + P(Z \geq 1.371) \\
&= 0.0054 + 0.0852 \\
&= 0.0906
\end{aligned}$$

(c) If $\mu = 17$:

$$\begin{aligned}
\text{Power} &= P(\bar{X} \leq 11.674|\mu = 17) + P(\bar{X} \geq 18.326|\mu = 17) \\
&= P(Z \leq \frac{11.674 - 17}{12/\sqrt{50}}) + P(Z \geq \frac{18.326 - 17}{12/\sqrt{50}}) \\
&= P(Z \leq -3.138) + P(Z \geq 0.781) \\
&= 0.0008 + 0.2173 \\
&= 0.2181
\end{aligned}$$

(d) If $\mu = 13$:

$$\begin{aligned}
\text{Power} &= P(\bar{X} \leq 11.674|\mu = 13) + P(\bar{X} \geq 18.326|\mu = 13) \\
&= P(Z \leq \frac{11.674 - 13}{12/\sqrt{50}}) + P(Z \geq \frac{18.326 - 13}{12/\sqrt{50}}) \\
&= P(Z \leq -0.781) + P(Z \geq 3.138) \\
&= 0.2173 + 0.0008 \\
&= 0.2181
\end{aligned}$$

21. (a) $P(\bar{X} > 10|\mu = 0) = P(Z > \frac{10-0}{100/\sqrt{50}}) = P(Z > 0.707) = 0.24$.
    (b) 0. A Type II error is possible only when the null hypothesis is false.
    (c) 0. A Type I error is possible only when the null hypothesis is true.
    (d) $P(\bar{X} > 10|\mu = 5) = P(Z > \frac{10-5}{100/\sqrt{50}}) = P(Z > 0.353) = 0.36$.
    (e) $P(\bar{X} > 10|\mu = 10) = P(Z > \frac{10-10}{100/\sqrt{50}}) = P(Z > 0) = 0.5$.

9.6.2 What Factors Affect the Power of the Test?

22. (a) An increase in the sample size will increase the power of the test.
    (b) A decrease in the significance level will decrease the power of the test.
    (c) An increase in the distance between the true mean and the hypothesized mean will increase the power of the test.
    (d) A decease in the population variance will increase the power of the test.

9.7 One-Sided Test or Two-Sided Test?

23. If we choose a one-sided alternative hypothesis, then the test will have greater power than its two-sided counterpart *if the true mean lies in the direction specified by the alternative hypothesis*. The downside of the one-sided alternative is that if the true mean lies in the opposite direction, then it will be impossible to detect this difference from the hypothesized value. If we choose a two-sided alternative hypothesis, then the test will be able to detect a difference from the hypothesized value in either direction.

24. This is poor statistical practice. We should not let the sample data influence our choice of hypothesis. If we are interested in a difference from the null hypothesis in either direction, then a two-sided alternative should be chosen at the outset.

Choosing a one-sided alternative because it fits the observed data is cheating. If we used the approach given in this question, we would be reporting a p-value that is smaller than it should be and our stated conclusions would be misleading.

9.8 Statistical Significance and Practical Significance

25. (a) If the p-value of a hypothesis test is 0.0000001, then there is very strong evidence that the results of the test have practical importance. *False. A very small p-value means that there is very strong evidence against the null hypothesis. Whether or not these results have any practical importance is a different question (one that cannot be answered with a hypothesis test).*
    (b) If the sample size is extremely large, then even tiny, meaningless differences from the hypothesized value will be found statistically significant. *True.*
    (c) If the results of a hypothesis test are reported, then there is no need to also report the associated confidence interval for the parameter. *False. It is good practice to also report a confidence interval for the parameter of interest, to give an idea of the size of the effect.*

9.9 The Relationship Between Hypothesis Tests and Confidence Intervals

26. (a) No. The p-value is less than 0.05, implying 50 would fall outside of the 95% confidence interval.
    (b) Yes. The p-value is greater than 0.01, implying 50 would be contained in the 99% confidence interval.

27. B (the p-value is less than 0.05). Since the 95% confidence interval for μ does not contain 0, a two-sided test of H_0: $\mu = 0$ would result in a p-value that is less than 0.05.

9.10 Hypothesis Tests for μ When σ is Unknown

28. (a) Reject H_0 if $|t| \geq 1.761$.
    (b) Reject H_0 if $t \leq -1.345$.
    (c) Reject H_0 if $t \geq 1.345$.
    (d) Reject H_0 if $|t| \geq 2.056$.
    (e) Reject H_0 if $t \leq -3.183$.

(f) Reject H_0 if $t \geq 1.038$.

29. (a) Using software: p-value $= 0.0508$. From the table: $0.05 < p\text{-value} < 0.10$.
(b) Using software: p-value $= 0.9745$. From the table: $0.95 < p\text{-value} < 0.975$.
(c) Using software: p-value $= 0.0254$. From the table: $0.025 < p\text{-value} < 0.05$.
(d) Using software p-value $= 0.0833$. From the table: $0.05 < p\text{-value} < 0.10$.
(e) Using software: p-value $= 0.0416$. From the table: $0.025 < p\text{-value} < 0.05$.
(f) Using software: p-value $= 0.9584$. From the table: $0.95 < p\text{-value} < 0.975$.

9.10.1 Examples of Hypothesis Tests Using the t Statistic

30. (a) The hypotheses are:

$H_0 : \mu = 11.0$ (This type of burrito has a true mean fat content of 11.0 g.)
$H_a : \mu \neq 11.0$ (This type of burrito has a true mean fat content that differs from 11.0 g.)

(b) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{1.45}{\sqrt{6}} = 0.59196$.

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{11.49 - 11.0}{0.59196} \\ &= 0.828 \end{aligned}$$

(c) Since the alternative hypothesis is two-sided, the p-value is double the area in the tail, beyond the observed value of the test statistic. In this case, the p-value is double the area to the right of 0.828 under a t distribution with $n - 1 = 6 - 1 = 5$ degrees of freedom (see Figure 9.3).

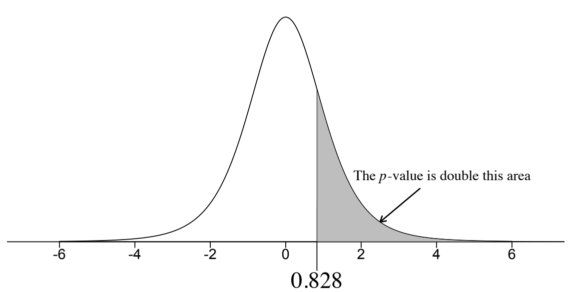


Figure 9.3: The p-value is double the area to the left of 0.828 under a t distribution with 5 degrees of freedom.

Using software, p-value $= 0.45$. Using a t table, we could find only a range: p-value > 0.40.

(d) Since the p-value is large (0.45), there is no evidence against the null hypothesis. There is no evidence that the true mean fat content in this type of bean burrito differs from the stated value of 11 grams. (The observed data is consistent with the true mean equalling 11 grams.) N.B. We cannot say that the true mean does equal 11.0 g, we can only that we have no evidence that it differs from 11 g.

(e) In order for the t procedure to be valid, we need the sample to be a simple random sample from the population of interest, and we need the population to be normally distributed. The sampling design was not clearly stated (we do not know how the sample was obtained), so we cannot make a truly informed decision about any possible biases that may have been introduced. While fast-food restaurants strive for consistency, it is likely that the amount of fat in a burrito tends to change a little from preparer to preparer, and possibly from restaurant to restaurant. But given the nature of fast food it is probably not too far off the mark to think of the sample as a random sample from all bean burritos from this fast-food chain.

The sample size is small, so the normality assumption is important. We were not given the data, but we were told that there were no outliers, which is somewhat comforting. Using the t procedure here is a little questionable, but still reasonable.

9.11 More on Assumptions

31. No, it would not be reasonable to use the t procedures on this raw data. This plot shows some right skewness and a large outlier. (Outliers and skewness negatively affect the t procedures.) The negative impact of skewness and/or outliers can be bad, especially when the sample size is not large. (The sample size here is 20, which is not very large.) There are other analysis options, such as using the t procedure on a *transformation* of the data (such as the log of the data values), or using a *distribution-free* procedure.

32. (a) The t test performs poorly when we are sampling from a strongly skewed distribution, especially if the sample size is small.

(b) If we use the t test when sampling from skewed distributions, then the *true* significance level may be quite different from the reported significance level, and the *true* p-value may be quite different from the stated p-value. In short, the reported results may be very misleading. This is especially true for small sample sizes.

9.12 Chapter Summary

9.12.1 Basic Calculations

33. (a) $t = \frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{15.5-20}{2/\sqrt{10}} = -7.12$. If the null hypothesis is true, the t test statistic will have a t distribution with $n-1 = 10-1 = 9$ degrees of freedom.
    (b) $t = \frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{15.5-20}{20/\sqrt{10}} = -0.712$. If the null hypothesis is true, the t test statistic will have a t distribution with $n-1 = 10-1 = 9$ degrees of freedom.
    (c) $Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{35.5-30}{2/\sqrt{40}} = 17.39$. If the null hypothesis is true, the Z test statistic will have the standard normal distribution. (Since we are sampling from a normally distributed population and σ is known.)
    (d) $t = \frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{45.5-40}{20/\sqrt{50}} = 1.94$. If the null hypothesis is true, the t test statistic will have a t distribution with $n-1 = 50-1 = 49$ degrees of freedom. (With these degrees of freedom, the t distribution will be similar to the standard normal distribution.)

34. (a) $t = \frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{35.5-30}{2/\sqrt{20}} = 12.30$. If the null hypothesis is true, the t test statistic will have a t distribution with $n-1 = 20-1 = 19$ degrees of freedom.
    (b) $Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{35.5-40}{20/\sqrt{25}} = -1.125$. If the null hypothesis is true, the Z test statistic will have the standard normal distribution. (Since we are sampling from a normally distributed population and σ is known.)
    (c) $t = \frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{65.5-30}{2/\sqrt{200}} = 251.0$. If the null hypothesis is true, the t test statistic will have a t distribution with $n-1 = 200-1 = 199$ degrees of freedom. (With these degrees of freedom, the t distribution is very close to the standard normal distribution.)
    (d) $t = \frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{32.1-32}{10/\sqrt{60000}} = 2.45$. If the null hypothesis is true, the t test statistic will have a t distribution with $n-1 = 60000-1 = 59999$ degrees of freedom. (With these degrees of freedom, the t distribution is very, very close to the standard normal distribution.)

35. (a) The null hypothesis is $\bar{X} = 50$. *False. We make hypotheses about parameters, not statistics.*
    (b) The sample size is 24. *False.* $n = 25$.
    (c) The area to the left of -4.1027 under a t distribution with 24 degrees of freedom is $2 \times 0.0004 = 0.0008$. *False. Twice the area to the left of* -4.1027 *is given in the output as the p-value (0.0004), so the area to the left of* -4.1027 *is 0.0002.*
    (d) If we used this data to test the null hypothesis that the population mean is 40.0, against a two-sided alternative hypothesis, the p-value would be less than 0.05. *True, since the 95% confidence interval does not contain 40.*

(e) The 95% margin of error is greater than 20.0. *False. The margin of error is the half-width of the interval:* $\frac{48.94322-46.80365}{2} = 1.069785$.

9.12.2 Concepts

36. For a one-sample Z test about a population mean μ to be valid, the sample must be a simple random sample from a normally distributed population. (We also need to know the true value of σ.) The assumption of a normally distributed population is not very important if the sample size is large.

37. For a one-sample t test about a population mean μ to be valid, the sample must be a simple random sample from a normally distributed population. The assumption of a normally distributed population is not very important if the sample size is large.

38. If the assumptions of a test are violated, then the reported significance level and p-value will be different from their "true" values, and the reported results may be misleading.

39. Both tests require the assumption of a normally distributed population, and for both tests this assumption becomes less important as the sample size increases. We use a Z test when the population standard deviation σ is known. We use a t test when the population standard deviation σ is unknown, and must be estimated by the sample standard deviation s.

40. The significance level, the hypothesized mean, and the choice of alternative hypothesis should *not* be based on the current sample's data. (One should be able to arrive at the appropriate choices before looking at the sample data.) The sample mean, the value of the test statistic, and the p-value do involve the current sample data (the test statistic and p-value also involve the appropriate hypotheses of the test).

41. (a) Here the normality assumption of the t test is true. If the null hypothesis is true, the t statistic will have a t distribution, and will therefore have a mean of 0.
    (b) Here the normality assumption of the t test is true. If the null hypothesis is true, the p-value will have a continuous uniform distribution on the interval (0,1). As such, the p-value will have a mean of 0.5.

42. (a) We are carrying out 25 independent tests, and the probability that we make a Type I error in any individual test is 0.05, so the number of Type I errors we make will have a binomial distribution with $n = 25$ and $p = 0.05$.

(b) Since the mean of a binomial random variable is np, on average we will make $np = 25 \cdot 0.05 = 1.25$ Type I errors.
(c) $P(\text{At least one Type I error}) = 1 - P(\text{No Type I errors}) = 1 - 0.95^{25} = 0.7226$.

43. B (0.0358). Since the hypothesized value of 29 is not contained in the 95% confidence interval, the p-value of the two-sided test would be less than 0.05. Since the hypothesized value of 29 *is* contained in the 99% confidence interval, the p-value of the two-sided test would be greater than 0.01. So, $0.01 < p\text{-value} < 0.05$. Of the 5 numbers listed, the only one that lies between 0.01 and 0.05 is 0.0358. (It is possible to verify that this is the correct p-value by carrying out the appropriate calculations.)

44. (a) $\bar{X} = -0.0377$ (the midpoint of the confidence interval).
(b) The margin of error is the half-width of the interval: $\frac{0.2656748-(-0.3412277)}{2} = 0.30345$.
(c) The p-value given in the output (0.8025) is double the area to the left of -0.2518 under a t distribution with 39 degrees of freedom. For an alternative of $\mu < 0$, the p-value is the area to the left of the test statistic, which is $\frac{0.8025}{2} = 0.40125$.
(d) Here we require the area to the *right* of the observed value of the test statistic, which is $1 - \frac{0.8025}{2} = 1 - 0.40125 = 0.59875$.
(e) A p-value of 0.008025 gives strong evidence against the null hypothesis. Here, there is strong evidence that the true population mean μ differs from 0.

45. All 3 values are possible values of μ. Although there is strong evidence against the null hypothesis, indicating that the population mean is likely less than 32, any value of μ is still possible.

46. Whether or not it is reasonable to use the t procedures is debatable in both cases.

The plot on the left shows strong right skewness (a clear violation of the normality assumption). Skewness negatively impacts the performance of the t procedures, but in this situation the large sample size of $n = 100$ would lessen the impact. In this situation, we would likely be better off doing an analysis involving a *transformation* of the data.

The plot on the right shows no clear violation of the normality assumption, but the sample size is very small. In order to use the t procedures when the sample size is this small, we need to be very confident that the normality assumption is reasonable. (Use of the t procedures is dubious when the small sample size is so small.)

47. (a) $SE(\bar{X}) = \frac{s}{\sqrt{40}} = 0.591$.
(b) There is very strong evidence (two-sided p-value = 0.00017) against the null

hypothesis. There is very strong evidence that the true population mean differs from 50. The point estimate of μ is 52.46, with an associated 95% confidence interval of (51.26, 53.66), giving strong evidence that the population mean is in fact greater than 50.

(c) 0.05. The significance level of a test is the probability of a Type I error (the probability of rejecting the null hypothesis, given the null hypothesis is true).

(d) The p-value is less than 0.05, so we would reject the null hypothesis at $\alpha = 0.05$. But if $\mu = 50$, the null hypothesis is true. We would have rejected a true null hypothesis, and thus we would have committed a Type I error.

(e) No. The plots give no indication of a violation of the normality assumption, so it is reasonable to carry out a t test.

48. (a) Sample B would have the largest p-value of this test (the least evidence that its population mean differs from 18).

(b) Sample C would have the smallest p-value of this test (the greatest evidence that its population mean differs from 18).

49. (a) No, the null hypothesis is not rejected (since the p-value is greater than the significance level).

(b) Yes, the null hypothesis is rejected (since the p-value is less than the significance level).

(c) C (but opinions may differ here). Since the p-values are nearly identical in the two tests, the amount of evidence against the null hypothesis is nearly identical.

50. A. If a Type I error is much, much worse than a Type II error, we should choose a very small value for α.

51. (a) A Type I error is rejecting the null hypothesis when it is true. If the null hypothesis is true, we will get tails (and reject the null hypothesis) with probability 0.5. The probability of a Type I error is thus 0.5.

(b) For it to be impossible to make a Type II error, it must be impossible to not reject the null hypothesis when it is false. This is true here only if the probability of tails is 1.

52. The null hypothesis H_0: $\bar{X} = 10$ is inappropriate, since we make hypotheses about parameters and not about statistics. The given p-value must be incorrect, as p-values are probabilities and cannot be greater than 1.

53. Since the one-sided p-value is 0.5, this implies $Z = 0$, which in turn implies $\bar{X} = \mu_0$. In this case, $\bar{X} = 50$.

54.

$$\begin{aligned}
P(\text{Type II error}) &= P(\text{Do not reject } H_0 \text{ when it is false}) \\
&= P(\bar{X} \leq 20.1 | \mu = 20.3) \\
&= P(Z \leq \frac{20.1 - 20.3}{0.5/\sqrt{100}}) \\
&= P(Z \leq -4.0) \\
&= 0.00003167
\end{aligned}$$

55. (a) $P(Z > 3) + P(Z < -3) = 0.0013 + 0.0013 = 0.0026$.
    (b) The probability of a Type I error will stay constant. (Due to the nature of her rejection rule, the probability of a Type I error for any sample size is what was calculated in 55a.)
    (c) 0. Since a Type I error is rejecting the null hypothesis when it is true, if the null hypothesis is false we cannot make a Type I error.
    (d) The power of her test will increase as the sample size increases.

56. B (the reported p-value will be less than it should be). The standard normal distribution has less area in the tails than the t distribution, and the two-sided p-value is double the area in the tail, beyond the observed value of the test statistic.

57. (a) All else being equal, the power of a test will increase as the population variance decreases. *True. With lower variability there is less uncertainty and a greater ability to detect deviations from the null hypothesis.*
    (b) If the significance level of a test is decreased (from 0.10 to 0.05, say), then the probability of a Type II error will increase. *True.*
    (c) The probability of a Type II error decreases the farther the true population mean is from the hypothesized mean. *True (for two-sided alternatives), and true in spirit for one-sided alternatives.*
    (d) If the null hypothesis is true, we cannot make a Type II error. *True.*

58. (a) Power $= 1 - P(\text{Type I error})$. *False.*
    (b) Power $= 1 - P(\text{Type II error})$. *True.*
    (c) $P(\text{Type I error}) = 1 - P(\text{Type II error})$. *False.*
    (d) All else being equal, the power of a test will decrease the closer the *true* value of the population mean is to the *hypothesized* value of the population mean. *True.*
    (e) All else being equal, the power of a test will increase as the sample size increases. *True.*

59. (a) If a null hypothesis is rejected at the 5% significance level, then it would also be rejected at the 10% significance level. *True. If we reject H_0 at $\alpha = 0.05$, this implies p-value ≤ 0.05, and we would also reject at $\alpha = 0.10$.*

(b) If a null hypothesis is rejected at $\alpha = 0.05$, then it would also be rejected at $\alpha = 0.01$. *False. If we reject H_0 at $\alpha = 0.05$, then p-value ≤ 0.05. We cannot tell if the p-value is less than 0.01.*

(c) If a null hypothesis is not rejected at the 5% significance level, then it will definitely not be rejected at the 10% significance level. *False. If we do not reject at $\alpha = 0.05$, we know that p-value > 0.05, and it may or may not be greater than 0.10.*

(d) The significance level of a test is $1 - p$-value. *False.*

(e) The significance level of a test is the probability the null hypothesis is true. *False.*

60. (a) If the p-value of a hypothesis test with a two-sided alternative is equal to exactly 0, we can be certain the null hypothesis is false. (This is a conceptual question—for all the tests we have done, it is not possible to get a p-value of exactly 0.) *True, since this would mean it was impossible to observe what was observed if the null hypothesis were true.*

(b) The p-value of a test is the probability of getting an insignificant result, given the null hypothesis is true. *False.*

(c) The p-value of a test is less than the significance level whenever the null hypothesis is false. *False.*

(d) The p-value of a test is the probability of rejecting the null hypothesis when it is false. *False.*

(e) If the p-value of a test is large (0.99, say), then we can be almost positive that the null hypothesis is true. *False.*

61. (a) For a two-sided alternative hypothesis, the p-value of the test will equal 1 if the sample mean equals the hypothesized mean. *True.*

(b) If we are sampling from a normally distributed population and conducting an ordinary t test, then if the null hypothesis is true the p-value will have a uniform distribution between 0 and 1. *True.*

(c) If the null hypothesis is true, then on average the p-value will equal 0.5. *True.*

(d) If the null hypothesis is false, then on average the p-value will equal 0.5. *False.*

9.12.3 Applications

62. (a) The boxplot and the normal QQ plot both show indications of right-skewness (which is typical of BMI data). This skewness is not a serious problem in this scenario, as the skewness is not very strong, and the sample size is fairly large (94). We can use the t procedures.

(b) The hypotheses are:

$$H_0 : \mu = 22.4 \quad \text{(The true mean BMI equals 22.4.)}$$
$$H_a : \mu \neq 22.4 \quad \text{(The true mean BMI differs from 22.4.)}$$

(c) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{5.4}{\sqrt{94}} = 0.5570.$

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{23.3 - 22.4}{0.5570} \\ &= 1.616 \end{aligned}$$

(d) Since the alternative hypothesis is two-sided, the p-value is double the area to the right of 1.616 under a t distribution with $n - 1 = 94 - 1 = 93$ degrees of freedom (see Figure 9.4). Using software, we can find that the p-value is 0.1095.

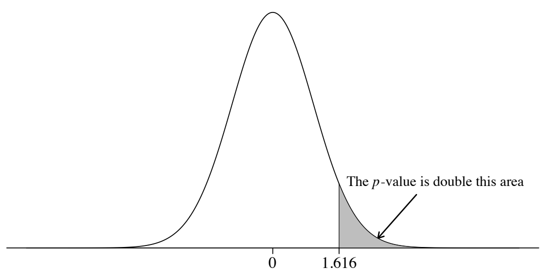


Figure 9.4: The p-value is double the area to the right of 1.616 under a t distribution with 93 degrees of freedom.

Using a table, we could find only a range of values: $0.10 < p\text{-value} < 0.20$.

(e) Because the p-value is not small (p-value $= 0.1095$), there is not strong evidence against the null hypothesis (The evidence against the null hypothesis would not be significant at $\alpha = 0.05$ or even $\alpha = 0.10$.) There is not strong evidence that the population mean BMI for 14 year-old boys in rural Texas differs from that of 14 year-old boys in the United States as a whole. (It is plausible that they are the same.)

(f) There is little or no evidence (p-value $= 0.1095$) against the null hypothesis that the true mean BMI for 14 year-old boys in rural Texas differs from that of 14 year-old boys in the United States as a whole. We can be 95% confident that the population mean BMI for 14 year-old boys in rural Texas lies between 22.2 and 24.4 kg/m^2. (Note that the 95% confidence interval captures 22.4, the hypothesized value of the population mean.)

The sample is drawn from a school district in North Central Texas, so very loosely speaking the population is all 14 year-old boys in north central Texas. However, there are some sampling biases that make this assessment not quite true:

- The sample was drawn from only 6 schools in the area. Students in these schools may differ in a meaningful way from those at other schools in the area (or from children that don't go to school).
- For students to participate, parents needed to provide informed consent. Students whose parents provide informed consent may differ in a meaningful way from those whose parents do not.
- Students needed to provide self-assent (the student had to agree to be part of the study). Students who agree to be part of a study like this may differ in a meaningful way from those who do not.
- Students who participated were given a T-shirt. This could introduce bias, as it is conceivable that there is a relationship between desire for the t-shirt and BMI.

63. (a) The plots look good, and show no indication of a violation of the normality assumption. There are no outliers, obvious skewness, or other problems. While we would prefer a larger sample size ($n = 11$ here), it is reasonable to use the t procedures.

(b) H_0: $\mu = 0$. (The true mean difference is 0. On average, the BR estimation method is equal to the true age at death.)
H_a: $\mu \neq 0$. (The true mean difference is not 0. On average, the BR estimation method does not equal the true age at death.)

(c) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{7.784}{\sqrt{11}} = 2.3470$.

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{11.073 - 0}{2.3470} \\ &= 4.718 \end{aligned}$$

(d) Since the alternative hypothesis is two-sided, the p-value is double the area to the right of 4.718 under a t distribution with $n - 1 = 11 - 1 = 10$ degrees of freedom. (see Figure 9.6).
Using software, we can find that the p-value is 0.0008. Using a table, we could find only a range: p-value < 0.001.

(e) There is very strong evidence against the null hypothesis (p-value $= 0.0008$). There is very strong evidence that the true mean difference between the BR estimate and the true age differs from 0.

(f) There is very strong evidence against the null hypothesis (p-value $= 0.0008$). There is very strong evidence that the true mean difference between the BR

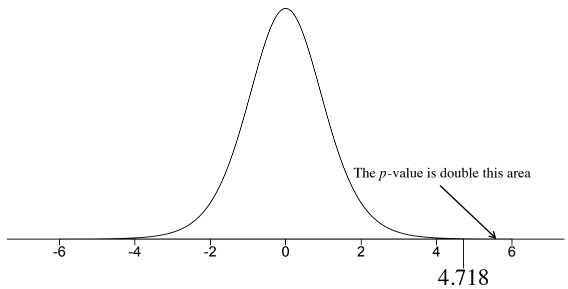


Figure 9.5: The p-value is double the area to the right of 4.718 under a t distribution with 10 degrees of freedom.

estimate and the true age differs from 0. We can be 95% confident that the true mean difference between the BR estimate and the true age lies between 5.84 years and 16.302 years. (Note that the hypothesized value of 0 falls *outiside* of the 95% interval.) It appears as though the true mean difference is in fact *greater* than the hypothesized value of 0, and therefore the BR estimate tends to be too large for individuals of this age.

The conclusions apply only to individuals that die at the age of 22 (but they would likely also hold for people of similar age). The sample is not truly a simple random sample from any population, and there may be strong sampling biases present, depending on the specific nature of the sample. While we are attempting to use these results to draw inferences about all people that die at 22, the BR method may have different performance for different subgroups. For example, it is possible that the BR aging method performs differently for different ethic groups. It is also conceivable that people who donate their bodies for educational or medical research differ in a meaningful way from those that do not. Experts in the field of dental aging would likely be able to provide guidance about the possible impacts of this type of bias.

64. (a) H_0: $\mu = 1$. (The true mean 2D:4D ratio equals 1.)
    H_a: $\mu \neq 1$. (The true mean 2D:4D ratio differs from 1.)

    (b) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.035}{\sqrt{135}} = 0.00301$.

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{0.994 - 1}{0.00301} \\ &= -1.992 \end{aligned}$$

    (c) Since the alternative hypothesis is two-sided, the p-value is double the area to the left of -1.992 under a t distribution with $n - 1 = 135 - 1 = 134$ degrees of freedom (see Figure 9.6).

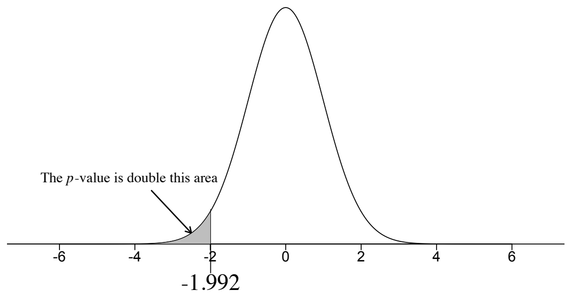


Figure 9.6: The p-value is double the area to the left of -1.992 under a t distribution with 134 degrees of freedom.

Using software, we can find that the p-value is 0.048. Using a table, we could find only a range: $0.02 < p\text{-value} < 0.05$.

(d) The p-value is less than the given significance level ($0.048 < 0.05$), so there is statistically significant evidence against the null hypothesis. There is strong evidence that the true mean 2D:4D ratio (on the left hand) for women of European descent at Western Washington University differs from 1.

(e) The p-value is less than the given significance level ($0.048 < 0.05$), so there is statistically significant evidence against the null hypothesis. There is strong evidence that the true mean 2D:4D ratio (on the left hand) for women of European descent at Western Washington University differs from 1. We can be 95% confident that the true mean 2D:4D ratio lies between 0.98804 and 0.99996. (Note that the hypothesized value of 1 falls *outiside* of the 95% interval.) It appears as though the true mean 2D:4D ratio is in fact *less* than the hypothesized value of 1. (Female students of European descent tended to have an index finger that is shorter than the ring finger.)

The sample was of female students of European descent in an introductory biology course at Western Washington University. So our conclusions apply to women of European descent enrolled in biology courses at this university, and to a lesser extent, women of European descent at the university. While we may look to studies like this to provide information about women of European descent in general, there are many possible biases to keep in mind, including:

- Students enrolled in an introductory biology course may tend to have different physical characteristics than other students.
- Students had to consent to being part of this study, and students who consent may possibly differ in a meaningful way from those that do not.
- Most importantly, women of European descent at this university may differ in a meaningful way from women of European descent in general.

While it is seemingly unlikely that these factors would have a major impact on the 2D:4D ratio, it is very possible that they would have a meaningful effect.

University students are not a random sample from the general population.

65. (a) The hypotheses are:

$H_0 : \mu = 30.0$ (This type of cheeseburger has a true mean protein content of 30 g.)
$H_a : \mu \neq 30.0$ (This type of cheeseburger has a true mean protein content that differs from 30 g.)

(b) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{2.72}{\sqrt{6}} = 1.1104.$

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{33.81 - 30.0}{1.1104} \\ &= 3.431 \end{aligned}$$

(c) Since the alternative hypothesis is two-sided, the p-value is double the area in the tail, beyond the observed value of the test statistic. In this case, the p-value is double the area to the right of 3.431 under a t distribution with $n - 1 = 6 - 1 = 5$ degrees of freedom (see Figure 9.7).

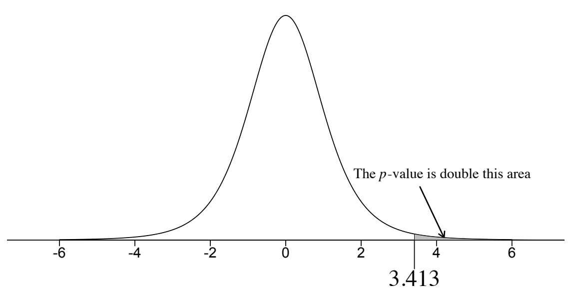


Figure 9.7: The p-value is double the area to the left of 3.431 under a t distribution with 5 degrees of freedom.

Using software, p-value = 0.019. Using a t table, we could find only a range: $0.01 < p\text{-value} < 0.02$.

(d) The p-value is less than the given significance level ($0.019 < 0.05$), so there is statistically significant evidence against the null hypothesis. There is statistically significant evidence that the true mean protein content in this type of cheeseburger differs from the stated value of 30 grams. Since the sample mean is greater than 30 (33.81), this implies there is strong evidence that the true mean protein content in this type of cheeseburger is in fact *greater* than 30 grams.

(e) In order for the t procedure to be valid, we need the sample to be a simple random sample from the population of interest, and we need the population to be normally distributed. The sampling design was not clearly stated (we do not know how the sample was obtained), so we cannot make a truly informed decision about any possible biases that may have been introduced. While fast-food restaurants strive for consistency, it is possible that the amount of protein in a cheeseburger tends to change a little from preparer to preparer, and from restaurant to restaurant. But given the nature of fast food it is probably not too far off the mark to think of the sample as a random sample from all cheeseburgers of this type from this fast-food chain.

The sample size is small, so the normality assumption is important. We were not given the data, but we were told that there were no outliers, which is somewhat comforting. Using the t procedure here is a little questionable, but still reasonable.

66. (a) H_0: $\mu = 5.0$ mg/kg dw. (The true mean lead concentration in caps of *Laccaria amethystina* in this area is 5.0 mg/kg of dry weight. (The true mean lead concentration is at the tolerable value.).)
H_a: $\mu > 5.0$ mg/kg dw. (The true mean lead concentration in caps of *Laccaria amethystina* in this area is greater than 5.0 mg/kg of dry weight. (The true mean lead concentration exceeds the tolerable value.).)
Note that the choice of alternative hypothesis is based on the problem at hand, and does not depend on the value of the sample mean.

(b) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.25}{\sqrt{21}} = 0.05455$.

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{0.94 - 5.0}{0.05455} \\ &= -74.421 \end{aligned}$$

(c) Since the alternative hypothesis is $\mu > 5.0$, the p-value is the area to the right of the observed value of the test statistic. In this case, the p-value is the area to the right of -74.421 under a t distribution with $n - 1 = 21 - 1 = 20$ degrees of freedom (see Figure 9.8).
For all intents and purposes, this area is 1. Using software: p-value $= 1 - 3.2 \times 10^{-26}$. Using a t table, we could find only a range: p-value > 0.40. (But we should have a pretty good idea that the p-value is near 1 just by looking at the value of the test statistic.)

(d) A p-value this large gives absolutely no evidence against the null hypothesis. There is no evidence whatsoever that the true mean lead concentration in this type of mushroom in this area is greater than the tolerable value of 5.0 mg/kg dw.

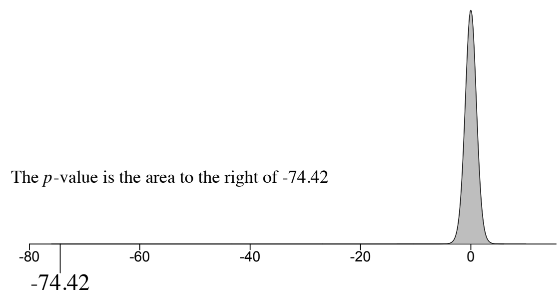


Figure 9.8: The p-value is the area to the right of -74.42 under a t distribution with 20 degrees of freedom. This area is very, very close to 1.

67. (a) No, there is no indication of a violation of the normality assumption. The boxplot looks roughly symmetric, and there is no systematic curvature or outliers evident in the normal QQ plot. A minor violation of the normality assumption would not be a big deal in this scenario, as the sample size is not small ($n = 51$). We can use the t procedures.

(b) The hypotheses are:

$H_0 : \mu = 1.0$ (The true mean speed of tandem running is 1.0 mm/s.)
$H_a : \mu \neq 1.0$ (The true mean speed of tandem running differs from 1.0 mm/s.)

(c) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{0.60}{\sqrt{51}} = 0.0840.$

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{1.80 - 1.0}{0.0840} \\ &= 9.522 \end{aligned}$$

(d) Since the alternative hypothesis is two-sided, the p-value is double the area to the right of 9.522 under a t distribution with $n - 1 = 51 - 1 = 50$ degrees of freedom (see Figure 9.9).

(e) There is very, very strong evidence against the null hypothesis (p-value $= 8.1 \times 10^{-13}$). There is very strong evidence that the true mean speed of tandem running (under the conditions of the study) differs from 1.0 mm/s.

(f) There is very, very strong evidence against the null hypothesis (p-value $= 8.1 \times 10^{-13}$). There is very strong evidence that the true mean speed of tandem running (under the conditions of the study) differs from 1.0 mm/s. We can be 95% confident that the true mean speed of tandem running (under the conditions of the study) lies between 1.63 mm/s and 1.97 mm/s. (Note

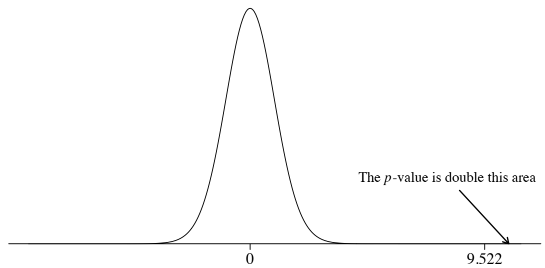


Figure 9.9: The p-value is double the area to the right of 9.522 under a t distribution with 50 degrees of freedom.

that the hypothesized value of 1.0 mm/s falls *outiside* of the 95% interval.) It appears as though the true mean speed of tandem running is in fact *greater* than the hypothesized value of 1.0 mm/s.

Although it wasn't perfectly clear from the article how the tandem running pairs were chosen, it is reasonable to think that these results apply to *Temnothorax albipennis* ants (with a young experienced ant leading a young inexperienced ant) under the conditions of the study. The true mean may be different if the conditions change, or if a different type of ant is studied.

9.12.4 Extra Practice Questions

68. (a) The normality assumption looks reasonable, and we can go ahead with the t procedures on this data.
    (b) The 90% interval is given by $1082 \pm 1.771 \times \frac{60}{\sqrt{14}}$, which works out to 1082 ± 28.399, or (1053.6, 1110.4).
    (c) H_0: $\mu = 920$, H_a: $\mu > 920$.
    (d) The t procedure is appropriate. Both the t and Z procedures assume a normally distributed population, which is a reasonable assumption here. Since the standard deviation is based on sample data (the population standard deviation is unknown), we would use the t procedure.
    (e) $t = \frac{1082-920}{60/\sqrt{14}} = 10.10$.
    (f) The p-value of the test is the area to the right of 10.10 under a t distribution with $n - 1 = 14 - 1 = 13$ degrees of freedom. Using software, the p-value can be found to be 8.0×10^{-8}. Using the t table, we could say only that the p-value is very close to 0 (depending on the specific table, we might say something like: p-value < 0.0005).
    (g) There is very strong evidence (one-sided p-value $= 8.0 \times 10^{-8}$) that the population mean calorie content is greater than what the company claims. A 90%

confidence interval for the population mean is found to be (1053.6, 1110.4), giving strong evidence that the true mean calorie content is more than 100 calories greater than what the company claims.
N.B. These conclusions rely heavily on the assumption of a simple random sample from the population of interest, and that assumption may not be reasonable. (For example, if all 14 bags were obtained from a new employee who was putting too much oil in the popcorn, that might seriously bias the results.)

69. (a) (ii) is correct (H_0: $\mu = 12.0, H_a$: $\mu > 12.0$). (Note once again that we make hypotheses about *parameters*, and never about *statistics*.)
(b) $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{16.7-12.0}{4.8/\sqrt{9}} = 2.9375$. The p-value is the area to the right of 2.9375 under a t distribution with $n - 1 = 9 - 1 = 8$ degrees of freedom. Using software: p-value $= 0.0188$. Using a t table, we could only give a range: $0.01 < p\text{-value} < 0.02$.
(c) Statement 2 (ii) is correct (and the rest of the statements are incorrect). Since the p-value is less than the given significance level of 0.05, the evidence against the null hypothesis is statistically significant at $\alpha = 0.05$. There is very strong evidence that the population mean arsenic level is greater than 12.0 ppm (the Environment Canada guideline).
(d) This might not be the most appropriate test for a number of reasons. An important reason is that the average arsenic concentration may not be the best summary of the data, since a high arsenic level in one school is very bad, regardless of whether it is offset by low levels in other schools. Also, we may wish to look at the hypotheses from the other perspective, and test to see if there is strong evidence that the population mean arsenic level is *less* than the Environment Canada guideline.
(e) We rejected a true null hypothesis, and therefore we made a Type I error.

70. (a) There is extremely strong evidence (p-value $= 7.7 \times 10^{-11}$) that the mean response time for this type of rat under these conditions differs from 1.25 seconds. The point estimate of the mean response time is 1.476, with an associated confidence interval of (1.42, 1.53). This gives strong evidence that the true mean response time is *greater* than 1.25 seconds.
(b) Although it is necessary from time to time, it is unwise to compare sample data to a "historical control", as it opens up a host of problems. Perhaps the rats in this experiment have slower response times due to characteristics of the lab in which the experiment took place, and not because of an effect of the drug. (The lab might have a poor measuring device, or low oxygen levels, or some other type of improper setup.) It is *much* better to conduct a case-control experiment, comparing two groups of rats under similar conditions.

71. (a) $n = 45$.
(b) 0.1622. Since the alternative hypothesis in the output is two-sided, the p-

value given in the output is double the area to the right of 0.9966 under a t distribution with 44 degrees of freedom. For the alternative H_a: $\mu > 250$, the p-value is the area to the right of 0.9966, so for this alternative: p-value $= \frac{0.3244}{2} = 0.1622$.

(c) 0.8378. Since the alternative hypothesis in the output is two-sided, the p-value given in the output is double the area to the right of 0.9966 under a t distribution with 44 degrees of freedom. For the alternative H_a: $\mu < 250$, the p-value is the area to the left of 0.9966, so for this alternative: p-value $= 1 - \frac{0.3244}{2} = 0.8378$.

(d) There is no evidence (p-value $= 0.32$) that the true mean weight of nuts in packages of this type differs from 250 grams.

(e) The third statement (iii) is a reasonable interpretation of the confidence interval, and the rest of the statements are incorrect.

72. (a) The appropriate hypotheses are given in the first option: H_0: $\mu = 0.50$, H_a: $\mu > 0.50$.

(b) $t = \frac{0.538-0.50}{0.14/\sqrt{150}} = 3.32$.

(c) The p-value is the area to the right 3.32 under a t distribution with $n - 1 = 150 - 1 = 149$ degrees of freedom. Using software: p-value $= 0.00056$. Using the t table: $0.0005 < p\text{-value} < 0.001$.

(d) Since the p-value is less than the given significance level ($0.00056 < 0.05$) there is statistically significant evidence against the null hypothesis. There is very strong evidence that the population mean trans-fat content exceeds 0.50 grams per serving.

(e) Yes, this hypothesis test would still be informative. The procedure assumes a normally distributed population, but this assumption becomes less important as the sample size increases. For a sample size of $n = 150$, the central limit theorem tells us that these methods should still work reasonably well. (But there may be better methods of analysis. For example, we might wish to use the t procedures on a transformation of the data.)

73. (a) A t statistic is appropriate, since the observations were roughly normally distributed and the standard deviation is based on sample data.

(b) $t = \frac{84.14-82.9}{0.12/\sqrt{100}} = 103.33$.

(c) The p-value is very close to 0. The p-value is twice the area to right of 103.33 under a t distribution with $100 - 1 = 99$ degrees of freedom. This area is very, very near zero. (While conceptually there is some non-zero area this far out in the tail, for all intents and purposes this area is 0.)

(d) Since the p-value is less than the given significance level of 0.01, we can say there is statistically significant evidence against the null hypothesis. There is extremely strong evidence that the population mean weight of bolts of this type exceeds 82.9 grams.

(e) We would have rejected a true null hypothesis, and therefore we would have made a Type I error.

74. (a) H_0: $\mu = 800$ (the true mean head injury rating for drivers is 800).
H_a: $\mu < 800$ (the true mean head injury rating for drivers is less than 800).
(b) $t = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{750-800}{50/\sqrt{24}} = -4.90$.
(c) The p-value is the area to the left of -4.90 under a t distribution with $n-1 = 24-1 = 23$ degrees of freedom. Using software: p-value $= 3.0 \times 10^{-5}$. Using the t table: p-value < 0.0005.
(d) There is very strong evidence (p-value $= 3.0 \times 10^{-5}$, significant at $\alpha = 0.05$) that the population mean driver head injury rating for this model of car under these conditions is less than 800.

75. (a) H_0: $\mu = 7$, H_a: $\mu \neq 7$. $t = \frac{6.8-7}{0.32/\sqrt{8}} = -1.768$. Using software: p-value $= 0.1204$. Using the t table: $0.10 < p$-value < 0.20. There is perhaps a tiny hint of evidence (p-value $= 0.12$) that the true mean fuel consumption under the driving conditions of the test differs from 7.0 l/100km, but it is not statistically significant at any of the usual levels.
(b) We need the sample to be a simple random sample from a normally distributed population. (Can we view the observed values as a random sample from a population of fuel consumption values for this model of car under these driving conditions?) The assumption of a random sample from the population could be violated here (for example, the manufacturer might have hand-picked the most recently tuned-up cars, or drivers might alter their driving habits in an effort to improve fuel efficiency). The normality assumption could be violated (the distribution of fuel consumption might be skewed, or have other types of non-normality.) If the assumptions are violated then the *stated* p-value will differ from the *true* p-value, and thus the stated conclusions might be misleading.

Chapter 10

Inference for Two Means

J.B.'s strongly suggested exercises: 4, 6, 7, 9, 10, 11, 17, 23, 26, 27, 30, 31, 33

10.1 Introduction

10.2 The Sampling Distribution of the Difference in Sample Means

1. (a) $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 = 15 - 8 = 7$.
   (b) $\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2^2}{4} + \frac{6^2}{4}} = \sqrt{10}$.
   (c) It is impossible to know the shape of the distribution without further information. We do not know the shape of the distributions from which we are sampling, and the sample sizes are not large enough for us to count on the central limit theorem to provide approximate normality.
   (d) $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 = 15 - 8 = 7$.
   (e) $\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2^2}{400} + \frac{6^2}{400}} = \sqrt{0.1}$.
   (f) Although we cannot know the exact shape of the distribution without further information, the sample sizes are large enough for the distribution of $\bar{X}_1 - \bar{X}_2$ to be approximately normal (due to the central limit theorem).

2. (a) Let T be a random variable representing Tom's time on the first run, and P be a random variable representing Pete's time on the first run. We are given that $T \sim N(4.612, 0.048)$ and $P \sim N(4.528, 0.044)$.

$T-P$ has a mean of $\mu_T - \mu_P = 4.612 - 4.528 = 0.084$ and a standard deviation of $\sqrt{\sigma_T^2 + \sigma_P^2} = \sqrt{0.048^2 + 0.044^2} = 0.0651$. Since both T and P are approximately normal, $T-P$ is approximately normal. $P(T > P) = P(T - P > 0) = P(Z > \dfrac{0 - 0.084}{0.0651}) = P(Z > -1.290) = 0.901$.

(b) Let $\bar{X}_T$ represent the mean of Tom's 3 times and $\bar{X}_P$ represent the mean of Pete's 3 times. $\bar{X}_T - \bar{X}_P$ has a mean of $\mu_T - \mu_P = 4.612 - 4.528 = 0.084$ and a standard deviation of $\sqrt{\frac{\sigma_T^2}{3} + \frac{\sigma_P^2}{3}} = \sqrt{\frac{0.048^2}{3} + \frac{0.044^2}{3}} = 0.03759$. $P(\bar{X}_T > \bar{X}_P) = P(\bar{X}_T - \bar{X}_P > 0) = P(Z > \frac{0-0.084}{0.03759}) = P(Z > -2.234) = 0.987$.

10.3 Hypothesis Tests and Confidence Intervals for Two Independent Samples (When σ_1 and σ_2 are known)

10.4 Hypothesis Tests and Confidence Intervals for $\mu_1 - \mu_2$ (When σ_1 and σ_2 are unknown)

10.4.1 Pooled-Variance t Tests and Confidence Intervals

3. (a) The point estimate of the difference in the population means ($\mu_1 - \mu_2$) is the value of the difference in the sample means: $\bar{X}_1 - \bar{X}_2 = 8.8 - 17.2 = -8.4$.

   (b) $s_p^2 = \frac{(10-1)1.42^2 + (5-1)2.61^2}{10+5-2} = 3.492$.

   (c) $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{3.492}\sqrt{1/10 + 1/5} = 1.0235$.

   (d) $-8.4 \pm 2.160 \times 1.0235$, which works out to -8.4 ± 2.211 or $(-10.61, -6.19)$.

   (e) H_0: $\mu_1 = \mu_2$, H_a: $\mu_1 \neq \mu_2$.
   $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{3.492}\sqrt{1/10 + 1/5} = 1.023523$. $t = \frac{8.8-17.2}{1.0235} = -8.21$.
   The p-value is double the area to the left of -8.21 under a t distribution with $10+5-2 = 13$ degrees of freedom. Using software: p-value $= 1.7 \times 10^{-6}$. Using a t table: p-value < 0.001. There is statistically significant evidence against the null hypothesis (p-value $= 1.7 \times 10^{-6}$) at $\alpha = 0.05$. There is statistically significant evidence that the population means are not equal.

4. (a) H_0: $\mu_1 = \mu_2$ (Vitamin D has no effect on parathyroid hormone levels).
   H_a: $\mu_1 \neq \mu_2$ (Vitamin D has an effect on parathyroid hormone levels).
   $s_p^2 = \frac{(14-1)37.5^2 + (12-1)34.6^2}{14+12-2} = 1310.417$
   $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{1310.417}\sqrt{1/14 + 1/12} = 14.24088$. $t = \frac{-9.0-(-1.6)}{14.24088} = -0.52$.
   The p-value is double the area to the left of -0.52 under a t distribution with $14 + 12 - 2 = 24$ degrees of freedom. Using software: p-value $= 0.61$. Using a

t table: p-value > 0.40.
There is no evidence (p-value $= 0.61$) that vitamin D has an effect on PTH levels (there is no evidence of a difference in population mean PTH levels between the groups).

(b) $-9.0-(-1.6)\pm 2.064\times 14.24088$, which works out to -7.4 ± 29.39, or $(-36.79, 21.99)$. We can be 95% confident that the difference in the population mean change in PTH lies between -36.79 pg/mL and 21.99 pg/mL. (It should come as no surprise that this interval contains 0, as the hypothesis test in part a) showed no evidence of a difference in means.)

(c) There is no evidence of an effect of vitamin D on PTH levels in the blood (p-value $= 0.61$). The point estimate of the difference in means (Vitamin D vs control) was -7.4 pg/mL, with a corresponding 95% confidence interval of $(-36.79, 21.99)$.

10.4.2 Welch (Unpooled Variance) t Tests and Confidence Intervals

5. (a) $\bar{X}_1 - \bar{X}_2 = 8.8 - 17.2 = -8.4$.

   (b) $SE_W(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{1.42^2}{10} + \frac{2.61^2}{5}} = 1.2506$. (Note that this is a little different from the standard error of the pooled-variance procedure.)

   (c) To construct the interval, we need to find $t_{0.025}$ with 5.221 degrees of freedom. Using software, we can find that with 5.221 degrees of freedom, $t_{0.025} = 2.538$. Using a t table, we would round the degrees of freedom down to 5, and find that $t_{0.025} = 2.571$.
   Using software, the interval is: $-8.4 \pm 2.538 \times 1.2506$, which works out to -8.4 ± 3.174 or $(-11.57, -5.23)$.
   Using a t table (with 5 DF), the interval is: $-8.4 \pm 2.571 \times 1.2506$, which works out to -8.4 ± 3.215 or $(-11.62, -5.18)$. (There is, of course, a little rounding error here.)

   (d) H_0: $\mu_1 = \mu_2$, H_a: $\mu_1 \neq \mu_2$.
   $SE_W(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{1.42^2}{10} + \frac{2.61^2}{5}} = 1.2506$. $t = \frac{8.8-17.2}{1.2506} = -6.72$.
   The p-value is double the area to the left of -6.72 under a t distribution with 5.221 degrees of freedom. Using software: p-value $= 0.0009$. Using a t table (with 5 DF): $0.001 < p$-value < 0.002. There is statistically significant evidence against the null hypothesis (p-value $= 0.0009$) at $\alpha = 0.05$. There is statistically significant evidence that the population means are not equal.

6. (a) $SE_W(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{18.7^2}{14} + \frac{1.7^2}{12}} = 5.0218$.

   (b) $t = \frac{-6.3-(-0.2)}{5.0218} = -1.215$

   (c) There is little or no evidence (p-value $= 0.245$) of an effect of vitamin D on phosphorous levels (little or no evidence of a difference in population mean phosphorous levels between the groups). The point estimate of the difference in change in phosphorous levels (Vitamin D vs control) was −6.1 mg/dL, with a corresponding 95% confidence interval of $(-16.9, 4.7)$.

10.4.3 Guidelines for Choosing the Appropriate Two-Sample t Procedure

7. Opinions differ a great deal on the choice between the Welch procedure and the pooled-variance t procedure. Both procedures assume normally distributed populations. The pooled-variance procedure also assumes that the populations have equal variances. If the *sample* variances are very different, then this gives some indication that the assumption of equal population variances is not reasonable, and the pooled-variance procedure may not be performing well. The effect of different population variances is exaggerated if the sample sizes are also very different. So if the sample variances are very different, and the sample sizes are very different, then the pooled-variance procedure is not an appropriate method of analysis. If the sample variances are similar, and the sample sizes are similar, then the choice between the Welch procedure and the pooled-variance t procedure is largely a matter of personal preference.

8. D, since the pooled-variance procedure performs poorly in this situation.

10.5 Paired-Difference Procedures

10.5.1 Paired-Difference t Tests and Confidence Intervals

9. (a) We should use a paired-difference procedure, since the before and after measurements represent *two measurements on the same pig.* (The before and after measurements do not represent independent samples.)

(b) The 9 After – Before measurements are:

$$\begin{aligned}
935 - 685 &= 250\\
700 - 545 &= 155\\
770 - 480 &= 290\\
640 - 475 &= 165\\
800 - 680 &= 120\\
955 - 685 &= 270\\
780 - 590 &= 190\\
790 - 600 &= 190\\
830 - 630 &= 200
\end{aligned}$$

(c) $\bar{X} = \frac{250+155+290+165+120+270+190+190+200}{9} = 203.333$ mL. $s = 56.1805$ mL.

(d) $SE(\bar{X}) = \frac{56.1805}{\sqrt{9}} = 18.7268$ mL.

(e) The boxplot and normal QQ plot look good. There are no obvious outliers, skewness, or other major problems. The sample size is quite small ($n = 9$), but it is not unreasonable to use the t procedures here.

(f) Using software, we can find that with 8 degrees of freedom, $t_{0.025} = 2.306004$. (If we must use a t table, we can find the rounded t value of 2.306.)

$$\begin{aligned}
&\bar{X} \pm t_{0.025} SE(\bar{X})\\
&203.333 \pm 2.306 \times 18.7268\\
&203.333 \pm 43.184
\end{aligned}$$

which works out to (160.15, 246.52).

We can be 95% confident that the true mean increase in leg volume, for this type of pig under the conditions of this experiment, lies between 160.15 mL and 246.52 mL.

(g) The hypotheses are:

$H_0 : \mu = 0$ (The true mean change in leg volume (under the conditions of this experiment) is 0.)

$H_a : \mu > 0$ (The true mean change in leg volume is greater than 0.)

$$\begin{aligned}
t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})}\\
&= \frac{203.333 - 0}{18.7268}\\
&= 10.8579
\end{aligned}$$

The p-value is the area to the right of 10.8579 under a t distribution with $n-1=9-1=8$ degrees of freedom. Using software, we can find: p-value $= 2.3\times10^{-6}$. Using a t table, we could find only a range: p-value < 0.0005.

A p-value that is this small gives very strong evidence against the null hypothesis. (The evidence would be statistically significant at any of the usual significance levels.) There is very strong evidence that, on average, pigs treated with a dose of venom and antivenom experience swelling in the affected area.

10. (a) The data should be plotted. Plotting the data helps us get a feel for what the numbers are telling us, and can help us to determine whether the normality assumption is reasonable.

    (b) The hypotheses are:

    $H_0: \mu = 0$ (The true mean change in discrimination accuracy is 0.)

    $H_a: \mu > 0$ (The true mean change in discrimination accuracy is greater than 0.)

    (In loose terms, the null hypothesis is that the short training program did not help people discriminate between genuine and fake smiles, and the alternative hypothesis is that the training program (on average) helped people to discriminate between genuine and fake smiles.)

    The sample mean fell in the direction given in the alternative hypothesis, so the one-sided p-value is half of the reported two-sided p-value. One-sided p-value $= \frac{0.0348}{2} = 0.0174$.

    (c) There is statistically significant evidence at $\alpha = 0.05$ (one-sided p-value $= 0.0174$) that the training program results in improved discrimination accuracy scores on average. The point estimate of the true mean change in discrimination accuracy is 0.618, with an associated 95% confidence interval of 0.048 to 1.188.

10.6 Investigating the Normality Assumption

10.7 Chapter Exercises

10.7.1 Basic Calculations

10.7.2 Concepts

11. $SE(\bar{X}_1 - \bar{X}_2)$ is the estimate of the standard deviation of the sampling distribution of $\bar{X}_1 - \bar{X}_2$. (It estimates $\sigma_{\bar{X}_1 - \bar{X}_2}$). It is a measure of the dispersion of $\bar{X}_1 - \bar{X}_2$ in repeated sampling.

12. No, this would not make sense. Statistical hypotheses always involve *parameters* and never *statistics.*

13. (a) Any hypothesized value outside of the 95% interval would be rejected at $\alpha = 0.05$. For example, H_0: $\mu_1 - \mu_2 = 42$ would be rejected.
    (b) Any hypothesized value within the 95% interval would not be rejected at $\alpha = 0.05$. For example, H_0: $\mu_1 - \mu_2 = 20$ would not be rejected.

14. Yes, of course $\bar{X}_1$ does not equal $\bar{X}_2$. The output shows that $\bar{X}_1 = -0.127$ and $\bar{X}_2 = 0.082$, so we *know* the sample means are not equal. However, in the test of H_0: $\mu_1 = \mu_2$, there is not any evidence against the null hypothesis (p-value $= 0.5763$).

15. (a) All else being equal, as $\mu_1 - \mu_2$ gets closer to 0, the power of the test will decrease (it will be harder to detect a difference if the difference is small).
    (b) All else being equal, as the sample size increases, the power of the test will increase.

16. (a) If the sample means are equal ($\bar{X}_1 = \bar{X}_2$), then $t = 0$.
    (b) The p-value would equal 1 if $t = 0$ (which occurs if $\bar{X}_1 = \bar{X}_2$). If the sample means are equal, then one could not possibly get a sample with less evidence against the null hypothesis, with a resulting p-value of 1.
    (c) The p-value would be equal to exactly 0 if the test statistic was infinite (which would only occur if the standard error was 0, implying no variance within each sample). In practice, this will never occur, although the p-value at times may be very, very, very *close* to 0.

17. (a) H_0: $\mu_A = \mu_C$ (II) would have the smallest p-value, since it shows the *greatest evidence against the null hypothesis.*

(b) H_0: $\mu_A = \mu_B$ (I) would have the largest p-value, since it shows the *least evidence against the null hypothesis.*

18. (a) The boxplots do not give any indication that the pooled-variance t procedure is inappropriate. (The pooled-variance t procedure is a reasonable choice.) There is one small outlier in Sample B, which may distort things a little bit, but it is not very extreme and it is not likely to have a major impact. The variances of the two samples are similar, so the equal variance assumption may not be too far off the mark.
    (b) No, there is not strong evidence (p-value $= 0.29$) of a difference in population means.
    (c) Yes. We can be 95% confident that $\mu_1 - \mu_2$ lies between -1.69 and 0.52. Since 10 falls far outside of this interval, there is very strong evidence that the difference in population means does not equal 10.
    (d) For the pooled variance procedure, $df = n1 + n2 - 2 = 43$ (from the output). This implies that $n_1 + n_2 = 45$.

19. (a) Output 1 is from the Welch procedure and Output 2 is from the pooled-variance procedure. (One can tell by looking at the degrees of freedom.)
    (b) The conclusions are nearly identical, so it matters little which procedure is used. The sample standard deviations are not given in the output, but they were likely similar.

20. (a) Of the given choices, the paired-difference procedure is the most appropriate one. We have two measurements on each individual, so in this situation it would not be appropriate to use Welch's t or the pooled-variance t (since they assume that the samples are independent). We need to use the paired-difference approach.
    (b) The fact that the placebo injection was given first, then the drug two days later, is a cause for concern. It is conceivable (and not a major stretch), to think that the individuals may be more nervous on the first day of the study, at the first injection. Perhaps by the second day, the participants were more comfortable with the study and more relaxed. This may very well influence the results. It would be much better to *randomize* the order in which they received each treatment (the drug or the placebo).

21. We should use the paired-difference approach, since there is a dependence within each pair of twins. (Twins are more similar than non-twins.)

22. Here we require $1.96\sqrt{\frac{3^2}{n} + \frac{3^2}{n}} \leq 0.5$. This implies that $1.96\sqrt{\frac{18}{n}} \leq 0.5 \implies n \geq (\frac{1.96\sqrt{18}}{0.5})^2 \implies n \geq 276.5$. We would need to sample at least 277 members from each population.

23. (a) When we use the pooled-variance t procedure, it is because we *know* the populations have the same variance. *False. We are assuming that the (unknown) population variances are equal.*
    (b) The pooled-variance t procedure works well, even when the population variances are a little different. This is especially true if the sample sizes are similar. *True.*
    (c) It would be most appropriate to use the Welch procedure instead of the pooled-variance t procedure if the sample variances are very different and the sample sizes are very different. *True.*
    (d) If the conclusions from the Welch procedure and the pooled-variance t procedure are very similar, then it does not matter much which procedure is used. *True.*
    (e) The Welch procedure is an exact procedure, as long as $\bar{X}_1 = \bar{X}_2$. *False. The Welch procedure is an approximate procedure.*

24. (a) If $\mu_1 = \mu_2$, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is approximately symmetric about 0 for large sample sizes. *True. The sampling distribution of $\bar{X}_1 - \bar{X}_2$ has a mean of $\mu_1 - \mu_2$, and is approximately normal for large sample sizes.*
    (b) If $\bar{X}_1 = \bar{X}_2$, the sampling distribution of $\mu_1 - \mu_2$ is approximately symmetric about 0 for large sample sizes. *False. $\mu_1 - \mu_2$ is a fixed quantity, and does not have a sampling distribution.*
    (c) $SE(\bar{X}_1 - \bar{X}_2)$ is the true standard deviation of the sampling distribution of $\bar{X}_1 - \bar{X}_2$. *False. It is the estimate of this quantity.*
    (d) The pooled-variance t procedures work well, even when the variances for the two populations are very different, as long as the sample sizes are very different as well. *False. The pooled-variance procedure performs poorly in this situation.*
    (e) Suppose we are interested in testing the null hypothesis $\mu_1 = \mu_2$, against a two-sided alternative. All else being equal, the greater the difference between μ_1 and μ_2, the greater the power of the test. *True. The greater the difference, the easier it is to detect that difference.*

25. (a) Suppose we are constructing a confidence interval for $\mu_1 - \mu_2$. All else being equal, the greater the difference between μ_1 and μ_2, the wider the interval. *False. The width does not depend on $\mu_1 - \mu_2$.*
    (b) Suppose we are constructing a confidence interval for $\mu_1 - \mu_2$. All else being equal, the greater the difference between $\bar{X}_1$ and $\bar{X}_2$, the wider the interval. *False. The width does not depend on $\bar{X}_1 - \bar{X}_2$. ($\bar{X}_1 - \bar{X}_2$ is the midpoint of the interval.)*
    (c) Suppose we are constructing a confidence interval for $\mu_1 - \mu_2$. All else being equal, the greater the sample sizes, the narrower the interval. *True.*
    (d) Suppose we wish to test H_0: $\mu_1 = \mu_2$. We obtain random samples from the respective populations, run the appropriate test, and find that the p-value is

0.00000032. We can be very confident that our results have important practical implications. *False. There is strong evidence of a difference in population means, but whether this is important in a practical sense is an entirely different question.*

(e) If we test H_0: $\mu_1 = \mu_2$ against a two-sided alternative and find a p-value of 0.32, then we know that $\mu_1 = \mu_2$. *False. We do not have any evidence of a difference in population means, but we do not know the population means, and we do not know if they are equal.*

10.7.3 Applications

26. (a) The t procedures should be reasonable here. The boxplots appear to show that the distributions are roughly symmetric, but there is one (small) outlier in the OE group. The normal QQ plots are very roughly linear, although the OE group shows hints of possible left skewness (mainly due to the one small outlier). Overall the plots look reasonable, and since the two-sample t procedures are quite robust to violations of the normality assumption, the t procedures will likely perform reasonably well in this scenario.

(b) The sample standard deviations are not that similar (0.37 vs 0.60), and the sample sizes are quite different (15 vs 51), so it is probably best to use the Welch (unpooled variance) approach. (Opinions may differ on this matter.)

(c) Using software, we can find that with 37.714 degrees of freedom $t_{0.025} = 2.0249$. (If we used the table or software with 37 degrees of freedom, $t_{0.025} = 2.026$.)

$$\begin{gathered}
\bar{X}_1 - \bar{X}_2 \pm t_{0.025} SE_W(\bar{X}_1 - \bar{X}_2) \\
1.14 - 1.80 \pm 2.0249 \cdot 0.1272 \\
-0.66 \pm 0.258
\end{gathered}$$

which works out to $(-0.92, -0.40)$.

(d) H_0: $\mu_1 = \mu_2$ (The true mean running speed for pairs led by an OE ant is equal to the true mean running speed for pairs led by a YE ant.)
H_a: $\mu_1 \neq \mu_2$ (The true mean running speed for pairs led by an OE ant is not equal to the true mean running speed for pairs led by a YE ant.)

(e)

$$\begin{aligned}
t &= \frac{\bar{X}_1 - \bar{X}_2}{SE_W(\bar{X}_1 - \bar{X}_2)} \\
&= \frac{1.14 - 1.80}{0.1272} \\
&= -5.19
\end{aligned}$$

(f) The p-value is double the area to the left of -5.19 under a t distribution with 37.714 degrees of freedom. Using software, we can find that the p-value is 7.5×10^{-6}. Using a table, we could find only a range for the p-value: p-value < 0.001. This is a very small p-value, implying very strong evidence against the null hypothesis.

(g) There is very strong evidence (p-value $= 7.5 \times 10^{-6}$) of a difference in the true mean speed of YE and OE leaders in tandem running for *Temnothorax albipennis* under the conditions of the study. The 95% confidence interval indicates that we can be 95% confident that the difference ($\mu_{OE} - \mu_{YE}$) in true mean speed lies between -0.92 and -0.40 mm/s. In short, there is very strong evidence that, on average, tandem running pairs led by an old and experienced ant move more slowly than pairs lead by a young and experienced ant.

Although it wasn't clear from the question how the tandem running pairs were chosen, it is reasonable to think that these results apply to *Temnothorax albipennis* ants under the conditions of the study. The means of tandem running speed may very well change if the conditions change, or if a different type of ant is studied.

27. (a) There are a variety of options, but we should (at least) plot boxplots and normal quantile-quantile plots. This would help us visualize possible differences between the groups, and investigate whether the normality assumption is reasonable. (The sample sizes are fairly large here (246 and 46), so the normality assumption is not overly important, but we should still investigate the shape of the distribution and ensure there are no major outliers.)

(b) The sample standard deviations are very similar here (0.035 vs 0.034), so the pooled-variance procedure would be a reasonable choice. The Welch (unpooled variance) approach would also be reasonable. (Opinions may differ on this matter.)

(c) There are $n_1 + n_2 - 2 = 246 + 46 - 2 = 290$ degrees of freedom. Using software or a t table, we can find that with 290 degrees of freedom $t_{0.025} = 1.968$.

$$\bar{X}_1 - \bar{X}_2 \pm t_{0.025} SE(\bar{X}_1 - \bar{X}_2)$$
$$0.994 - 0.963 \pm 1.968 \cdot 0.005598$$
$$0.031 \pm 0.0110$$

which works out to $(0.020, 0.042)$.

(d) H_0: $\mu_1 = \mu_2$ (The true mean 2D:4D ratio for white women is equal to the true mean 2D:4D ratio for black women.)
H_a: $\mu_1 \neq \mu_2$ (The true mean 2D:4D ratio for white women is not equal to the true mean 2D:4D ratio for black women.)

(e)

$$\begin{aligned} t &= \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)} \\ &= \frac{0.994 - 0.963}{0.005598} \\ &= 5.538 \end{aligned}$$

(f) The p-value is double the area to the right of 5.538 under a t distribution with 290 degrees of freedom. Using software, we can find that the p-value is 6.8×10^{-8}. Using a table, we could find only a range for the p-value: p-value < 0.001. This is a very small p-value, implying very strong evidence against the null hypothesis.

(g) There is very strong evidence (p-value $= 6.8 \times 10^{-8}$) of a difference in the true mean 2D:4D ratio for white and black women in Manchester. The 95% confidence interval indicates that we can be 95% confident that the difference in true mean 2D:4D ratio ($\mu_W - \mu_B$) lies between 0.020 and 0.042. In short, there is very strong evidence that, on average, white women have a greater mean 2D:4D ratio than that of black women.

The women were drawn from people attending a sexual health clinic in Manchester, UK. It is reasonable to think that these results would apply to women who attend this clinic. Extrapolating to an even larger population is somewhat dubious, but it is likely reasonable to think that these results apply to white and black women in the Manchester, UK area. Although these results give some indication of the difference in this ratio for white and black women in general, generalizing beyond the area from which the sample was drawn is always dubious, and may lead to incorrect conclusions.

28. (a) There are a variety of options, but we should (at least) plot boxplots and normal quantile-quantile plots. This would help us visualize possible differences between the groups, and investigate whether the normality assumption is reasonable. (The sample sizes are small here (8 and 7) so the normality assumption is important. Also, outliers could have a strong effect on the results.)

(b) The sample standard deviations are very different here (25.4 vs 7.9), so the Welch (unpooled variance) approach would likely be the best option. (Opinions may differ on this matter.)

(c) The degrees of freedom are given as 8.512. Using software, we can find that with 8.512 degrees of freedom $t_{0.025} = 2.2821$. (If we had to use a t table with

8 degrees of freedom, $t_{0.025} = 2.306$.

$$\bar{X}_1 - \bar{X}_2 \pm t_{0.025} SE(\bar{X}_1 - \bar{X}_2)$$
$$48.9 - 23.8 \pm 2.2821 \cdot 9.4637$$
$$25.1 \pm 21.597$$

which works out to $(3.5, 46.7)$. (Had we used the table value of $t_{0.025} = 2.306$, we'd have found an interval of (3.3,46.9).)

(d) H_0: $\mu_1 = \mu_2$ (The true mean number of lever presses for the restricted food group is equal to the true mean number of lever presses for the sated group.) H_a: $\mu_1 \neq \mu_2$ (The true mean number of lever presses for the restricted food group is not equal to the true mean number of lever presses for the sated group.)

(e)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)} = \frac{48.9 - 23.8}{9.4637} = 2.652$$

(f) The p-value is double the area to the right of 2.652 under a t distribution with 8.512 degrees of freedom. Using software, we can find that the p-value is 0.0276. Using a table, we could find only a range for the p-value: $0.02 < p\text{-value} < 0.05$. There is strong, but not overwhelming evidence against the null hypothesis. The evidence against the null hypothesis would be significant at the commonly chosen significance level of 0.05.

(g) There is strong evidence (p-value $= 0.0276$) of a difference in the true mean number of lever presses under the conditions of this experiment. The 95% confidence interval indicates that we can be 95% confident that the difference in true mean number of lever presses ($\mu_{FR} - \mu_S$) lies between 3.5 and 46.7. In short, there is strong evidence that, on average, the food restricted group presses the heroin lever more often.

The conclusions apply to the type of rat used in the study, under the conditions of the experiment. Using a different type of rat (or another animal), or changing the conditions of the study, could easily change the means a great deal.

29. (a) There are a variety of options, but we should (at least) plot boxplots and normal quantile-quantile plots. This would help us visualize possible differences between the groups, and investigate whether the normality assumption is reasonable. (The sample sizes are not small here (66 and 28), so the normality assumption is not overly important, but we should still investigate the shape of the distribution.)

(b) The sample standard deviations are very similar here (1.50 and 1.64), so the pooled-variance procedure would be a reasonable choice. The Welch (unpooled variance) approach would also be reasonable. (Opinions may differ on this matter.)

(c) There are $n_1 + n_2 - 2 = 66 + 28 - 2 = 92$ degrees of freedom. Using software, we can find that with 92 degrees of freedom $t_{0.025} = 1.9861$. (If we had to use a t table, we may have to look up 90 DF, and find $t_{0.025} = 1.987$.)

$$\bar{X}_1 - \bar{X}_2 \pm t_{0.025} SE(\bar{X}_1 - \bar{X}_2)$$
$$95.71 - 95.41 \pm 1.9861 \cdot 0.3479$$
$$0.30 \pm 0.691$$

which works out to $(-0.39, 0.99)$. (Had we used $t_{0.025} = 1.987$ from the t table, we'd have gotten the same interval to 2 decimal places.)

(d) H_0: $\mu_1 = \mu_2$ (The true mean 3D symmetry for men is equal to the true mean 3D symmetry for women.)

H_a: $\mu_1 \neq \mu_2$ (The true mean 3D symmetry for men is not equal to the true mean 3D symmetry for women.)

(e)

$$\begin{aligned} t &= \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)} \\ &= \frac{95.71 - 95.41}{0.3479} \\ &= 0.862 \end{aligned}$$

(f) The p-value is double the area to the right of 0.862 under a t distribution with 92 degrees of freedom. Using software, we can find that the p-value is 0.39. Using a table, we could find only a range for the p-value: p-value > 0.20. This p-value is large, yielding no evidence against the null hypothesis.

(g) There is no evidence (p-value $= 0.39$) of a difference in the true mean 3D symmetry for men and women. The 95% confidence interval indicates that we can be 95% confident that the difference in true mean 3D symmetry ($\mu_M - \mu_W$) lies between -0.39 and 0.99. There is no evidence that, on average, men have different mean 3D symmetry than that of women.

The people in the sample were drawn from healthy Italian Caucasians between 31 and 40 years of age. Depending on the sampling design, it is likely reasonable to think that these results apply to healthy Italian Caucasians between 31 and 40 years old.

30. (a) There are a variety of options, but we should (at least) plot boxplots and normal quantile-quantile plots. This would help us visualize possible differences between the groups, and investigate whether the normality assumption is reasonable. (The sample sizes are not small here (47 and 27), so the normality

assumption is not overly important, but we should still investigate the shape of the distribution.)

(b) The sample standard deviations are similar here (33.53 and 37.55), so the pooled-variance procedure would be a reasonable choice. The Welch (unpooled variance) approach would also be reasonable.

(c) There are $n_1 + n_2 - 2 = 47 + 27 - 2 = 72$ degrees of freedom. Using software, we can find that with 72 degrees of freedom $t_{0.025} = 1.9935$. (If we had to use a t table, we may have to look up 70 DF, and find $t_{0.025} = 1.994$.)

$$\begin{gathered}\bar{X}_1 - \bar{X}_2 \pm t_{0.025} SE(\bar{X}_1 - \bar{X}_2) \\ 256.12 - 245.82 \pm 1.9935 \cdot 8.4603 \\ 10.3 \pm 16.867\end{gathered}$$

which works out to $(-6.6, 27.2)$. (Had we used $t_{0.025} = 1.994$ from the t table, we'd have gotten the same margin of error and interval to 2 decimal places.)

(d) H_0: $\mu_1 = \mu_2$ (The true mean of the BEST index for psychopaths is equal to the true mean of the BEST index for nonpsychopaths.)
H_a: $\mu_1 \neq \mu_2$ (The true mean of the BEST index for psychopaths is not equal to the true mean of the BEST index for nonpsychopaths.)

(e)

$$\begin{aligned} t &= \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)} \\ &= \frac{256.12 - 245.82}{8.4603} \\ &= 1.217 \end{aligned}$$

(f) The p-value is double the area to the right of 1.217 under a t distribution with 72 degrees of freedom. Using software, we can find that the p-value is 0.2274. Using a table, we could find only a range for the p-value: p-value > 0.20. This p-value is large, yielding little or no evidence against the null hypothesis.

(g) There is no evidence (p-value $= 0.22$) of a difference in the true mean BEST index of psychopaths and nonpsychopaths. The 95% confidence interval indicates that we can be 95% confident that the difference in true mean BEST index ($\mu_P - \mu_{NP}$) lies between -6.6 and 27.2. There is no evidence that, on average, psychopaths have a different mean BEST index than that of nonpsychopaths.

The people in the sample were males drawn from an inpatient psychiatric treatment facility in Holland. Technically the conclusions apply only to male psychiatric patients from this facility. (The results give some indication of the reality of the situation for male inpatient psychiatric patients in general, but generalizing to a population outside of the one sampled is always dubious.)

31. (a) $SE(\bar{X}) = \frac{s}{\sqrt{n}} = \frac{3.7}{\sqrt{11}} = 1.11559$.
Using software, we can find that with 10 degrees of freedom, $t_{0.025} = 2.228139$. (If we must use a t table, we can find the rounded t value of 2.228.) The endpoints of a 95% confidence interval for μ are given by:

$$\begin{aligned} &\bar{X} \pm t_{0.025} SE(\bar{X}) \\ &3.3 \pm 2.228 \times 1.11559 \\ &3.3 \pm 2.486 \end{aligned}$$

which works out to (0.814, 5.786).

We can be 95% confident that, when this type of pig is Tasered, the true mean change in mean corpuscular volume lies between 0.814 fL and 5.786 fL.

(b) The hypotheses are:

$H_0 : \mu = 0$ (The true mean change in mean corpuscular volume is 0.)
$H_a : \mu \neq 0$ (The true mean change in mean corpuscular volume differs from 0.)

$$\begin{aligned} t &= \frac{\bar{X} - \mu_0}{SE(\bar{X})} \\ &= \frac{3.3 - 0}{1.11559} \\ &= 2.958 \end{aligned}$$

The p-value is double the area to the right of 2.958 under a t distribution with $n-1 = 11-1 = 10$ degrees of freedom. Using software, we can find: p-value $=$ 0.0143. Using a t table, we could find only a range: $0.01 < p\text{-value} < 0.02$.

A p-value of this size gives moderately strong evidence against the null hypothesis. (The evidence would be statistically significant at the commonly chosen significance level of 0.05.) There is moderately strong evidence that, on average, Tasered pigs of this type experience an increase in their mean corpuscular volume.

10.7.4 Extra Practice Questions

32. (a) $3468-3200 \pm 1.684 \times 240.8803$, which works out to 268 ± 405.6 or $(-137.6, 673.6)$.
(b) ii) is the most appropriate conclusion.

33. (a) This is an experiment (since the researchers imposed a treatment, the exercise wheel, on the rats).

(b) H_0: $\mu_1 = \mu_2$ (the tumour size is the same on average for both groups).
H_a: $\mu_1 < \mu_2$ (the exercise group has a lower mean tumour size). (Here the researchers are only interested if the wheels tend to *reduce* tumour size, but one could certainly make an argument for a two-sided alternative hypothesis).
$s_p^2 = \frac{(10-1)0.08+(20-1)0.12}{10+20-2} = 0.10714$.
$SE(\bar{X}_1 - \bar{X}_2) = \sqrt{0.10714}\sqrt{1/10 + 1/20} = 0.12677$.
$t = \frac{1.2-1.5}{0.12677} = -2.366$. The p-value is the area to the left of -2.366 under a t distribution with 28 degrees of freedom. By computer: p-value $= 0.0126$. By the table: $0.01 < p$-value < 0.025.

Here we do not have a given value of α, but we should still be able to say something reasonable. There is strong evidence (p-value $= 0.0126$) of a difference in population mean tumour size. There is strong evidence that those rats in the exercise group have a *lower* population mean size of tumour.

(c) Yes, absolutely. A p-value that small gives very, very strong evidence against the null hypothesis. Since this was a designed experiment, we can say that there is evidence of a *causal* relationship (the exercised wheels caused the observed difference).

(d) $1.2-1.5\pm1.701\times0.12677$, which works out to -0.3 ± 0.2156 or $(-0.516, -0.084)$. We can be 90% confident that the true difference in mean tumour size lies between -0.516 and -0.084.

(e) H_0: $\mu_1 = \mu_2$ (the tumour size is the same on average for both groups).
H_a: $\mu_1 < \mu_2$ (the exercise group has a lower mean tumour size). (Here the researchers are only interested if the wheels tend to *reduce* tumour size, but one could certainly make an argument for a two-sided alternative hypothesis).
$SE_W(\bar{X}_1 - \bar{X}_2) = \sqrt{0.08/10 + 0.12/20} = 0.1183$.
$t = \frac{1.2-1.5}{0.1183} = -2.535$. The p-value is the area to the left of -2.535 under a t distribution with 21.764 degrees of freedom.
By computer: p-value $= 0.0095$. By the table (with 20 DF): $0.005 < p$-value < 0.01.

Here we do not have a given value of α, but we should still be able to say something reasonable. There is strong evidence (p-value $= 0.0095$) of a difference in population mean tumour size. There is strong evidence that those rats in the exercise group have a *lower* population mean size of tumour.

(f) By computer (with 21.764 DF): $1.2 - 1.5 \pm 1.718 \times 0.1183$, which works out to -0.3 ± 0.2032 or $(-0.503, -0.097)$.
By the table (with 21 DF): $1.2 - 1.5 \pm 1.721 \times 0.1183$, which works out to -0.3 ± 0.2036 or $(-0.504, -0.096)$.

(g) The choice of appropriate procedure here would be debatable. There is not a large difference between the variances, so the pooled-variance procedure would not be a terrible choice. But one could make an argument for each one, so the choice comes down largely to a matter of personal preference.

34. (a) The last statement (v) is the best conclusion. The other statements are not reasonable.
    (b) We can be 95% confident that the difference in population mean cholesterol levels between males and females ($\mu_M - \mu_F$, for students at this university) lies between -21.9 and 17.1 mg/dl.

35. (a) $s_p^2 = \frac{(12-1)16+(6-1)9}{12+6-2} = 13.8125$.
    $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{13.8125}\sqrt{1/12 + 1/6} = 1.8583$.
    A 95% confidence interval is $63 - 57 \pm 2.120 \times 1.8583$, which works out to 6 ± 3.94 or (2.06, 9.94). We can be 95% confident that the difference in population mean contaminant levels ($\mu_A - \mu_B$) lies between 2.06 and 9.94.
    (b) H_0: $\mu_1 = \mu_2$ (the mean contaminant level is the same for both groups).
    H_a: $\mu_1 \neq \mu_2$ (there is a difference in the mean contaminant level between the groups).
    $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{13.8125}\sqrt{1/12 + 1/6} = 1.8583$.
    $t = \frac{63-57}{1.8583} = 3.23$. The p-value is double the area to the right of 3.23 under a t distribution with 16 degrees of freedom. By computer: p-value $= 0.0053$. By the table: $0.002 < p\text{-value} < 0.010$.
    (c) There is strong evidence (p-value $= 0.0053$) of a difference in populate mean contaminant levels. The point estimate of the difference in means is 6, with a corresponding 95% confidence interval of $(2.06, 9.94)$. It appears as though process B results in a lower contaminant level on average.

36. (a) H_0: $\mu_1 = \mu_2$ (the population mean "fear of negative evaluation" is the same for bulimics and non-bulimics).
    H_a: $\mu_1 \neq \mu_2$ (the population mean "fear of negative evaluation" is different between bulimics and non-bulimics).
    (b) There is significant evidence against the null hypothesis. There is significant evidence of a difference in population mean "fear of negative evaluation" score. It appears as though bulimics tend to have a higher fear of negative evaluation score.
    (c) We can be 95% confident that the difference in population mean "fear of evaluation" scores lies between 0.94 and 8.80.

37. (a) No, these plots give no indication of a violation of the normality assumption. Although the sample sizes are small, the use of the t procedures is reasonable here.
    (b) $s_p^2 = \frac{(12-1)0.65^2+(16-1)0.42^2}{12+16-2} = 0.2805$.
    (c) H_0: $\mu_1 = \mu_2$ (The population mean resistance is the same for both types of resistor.)
    H_a: $\mu_1 \neq \mu_2$ (The population mean resistance differs between the two types

of resistor.)

(d) $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{0.2805}\sqrt{1/12 + 1/16} = 0.2023$.
$t = \frac{11.6-12.2}{0.2023} = -2.967$.

(e) The p-value is double the area to the left of -2.967 under a t distribution with 26 degrees of freedom. By computer: p-value $= 0.0064$. By the table: $0.002 <$ p-value < 0.010. There is significant evidence at the 5% significance level (p-value < 0.05) of a difference in population mean resistance. It appears as though the resistors from Manufacturer B have a greater mean resistance.

(f) $11.60-12.20 \pm 2.056 \times 0.2023$, which works out to -0.6 ± 0.416 or $(-1.016, -0.184)$.

(g) We can be 95% confident that the difference in the population mean resistance $(\mu_A - \mu_B)$ lies between -1.052 and -0.148.

(h) The output tells us that double the area to the left of -2.7904 under the t distribution is 0.01222. Since for the one-sided alternative H_a: $\mu_A < \mu_B$ we require the area to the left of -2.7904, the p-value is $\frac{0.01222}{2} = 0.0061$.

38. (a) $s_p^2 = \frac{(80-1)23.3^2+(75-1)19.8^2}{80+75-2} = 469.9299$.
$SE(\bar{X}_1 - \bar{X}_2) = \sqrt{469.9299}\sqrt{1/80 + 1/75} = 3.4842$.
A 95% confidence interval is given by $664.2 - 624.3 \pm 1.976 \times 3.4842$, which works out to 39.9 ± 6.88 or $(33.02, 46.78)$. We can be 95% confident that the difference in population mean yield strength $(\mu_{New} - \mu_{Standard})$ lies between 33.02 MPa and 46.78 MPa.

(b) H_0: $\mu_1 = \mu_2$ (the population mean yield strength is the same for both methods).
H_a: $\mu_1 \neq \mu_2$ (the population mean yield strength differs between the methods).
$SE(\bar{X}_1 - \bar{X}_2) = \sqrt{469.9299}\sqrt{1/80 + 1/75} = 3.4842$.
$t = \frac{664.2-624.3}{3.4842} = 11.45$.
The p-value is double the area to the right of 11.45 under a t distribution with $80 + 75 - 3$ degrees of freedom. Using software: p-value $= 2.6 \times 10^{-22}$. Using a t table: We can say that the p-value is close to 0 (p-value < 0.001).
There is extremely strong evidence (p-value $= 2.6 \times 10^{-22}$) of a difference in population mean yield strength between the two procedures. In fact, there is very strong evidence that the heat treated method results in a greater mean yield strength.

39. (a) There is strong evidence (p-value $= 0.028$) of a difference on average between the stated and actual weight of the chickens. The point estimate of the difference (Stated – Actual) is 22.8, with an associated 95% confidence interval of 4.1 to 41.5 grams. In loose terms, it appears as though the actual weight tends to be lower than the stated weight.

(b) If the normality assumption is violated then our reported results may be misleading. What we are *stating* is a 95% confidence interval will have a *true* confidence level that differs from 95%. And our reported p-value may be very different from the "true" p-value.

40. (a) Here we have 2 measurements on each gas sample, and thus these measurements cannot be considered independent. We need to use a paired-difference procedure.
    (b) We could try plotting a normal quantile-quantile plot of the 6 differences (but with only 6 points, the plot may not be very informative). It will be a little sketchy using the t procedures on such a small sample.
    (c) The 6 differences (Old − New) are $0.03, 0.10, -0.12, 0.05, -0.08, 0.08$. These 6 differences have a mean of $\bar{X} = 0.01$ and a standard deviation of $s = 0.08944$. $SE(\bar{X}) = \frac{0.08944}{\sqrt{6}} = 0.0365$. A 95% confidence interval for the population mean difference (Old − New) is $0.01 \pm 2.571 \times 0.0365$, which works out to 0.01 ± 0.094 or $(-0.084, 0.104)$. We can be 95% confident that the population mean difference (Old method − New method) lies between -0.084 and 0.104.
    (d) H_0: $\mu = 0$ (There is no difference between the two methods on average).
    H_a: $\mu \neq 0$ (There is a difference between the two methods on average).
    $t = \frac{0.01-0}{0.0365} = 0.274$.
    The p-value is double the area to the right of 0.274 under a t distribution with 5 degrees of freedom. By computer: p-value $= 0.80$. By the table: p-value > 0.40. There is no evidence (p-value $= 0.80$) that the population mean difference differs from 0. We have no evidence of a difference between the viscosity measurements on average.

41. (a) If we take the differences as After − Before:
    H_0: $\mu = 0$ (there is no change in reaction time on average).
    H_a: $\mu > 0$ (there is an increase in reaction time on average).
    The 4 differences are: $0.37, 0.25, 0.42, 0.06$. These values have a mean of $\bar{X} = 0.275$ and a standard deviation of $s = 0.1601$. $SE(\bar{X}) = \frac{0.1601}{\sqrt{4}} = 0.0801$. $t = \frac{0.275-0}{0.0801} = 3.44$. The p-value is the area to the right of 3.44 under a t distribution with 3 degrees of freedom. Using software: p-value $= 0.021$. Using a t table: $0.01 < p\text{-value} < 0.025$.
    There is significant evidence at at the 5% level that, on average, alcohol increases reaction times (under the conditions of this experiment).
    (b) $0.275 \pm 3.182 \times 0.0801$, which works out to 0.275 ± 0.255 or (0.02, 0.53). We can be 95% confident that the true mean increase in reaction times, under the conditions of this experiment, lies between 0.02 and 0.53.
    (c) Setting up the experiment in this way serves to reduce the variability in the experiment. By taking two measurements on the same person, we reduce the overall variability, and make it easier to isolate the effect of the alcohol.
    (d) Before-and-after experiments are not without their problems. Here it is conceivable that the students would be more comfortable with the testing procedure after going through the process once. This may serve to naturally

decrease (or possibly increase) reactions times. It is better to randomly assign which treatment (alcohol/sober) comes first. But here, in a practical setting, it is much easier to have the sober times be the "before" times.

Chapter 11

Inference for Proportions

J.B.'s strongly suggested exercises: 2, 5, 6, 8, 10, 12, 16, 17, 19, 20, 24, 25, 26, 27

11.1 Introduction

11.2 The Sampling Distribution of $\hat{p}$

1. $E(\hat{p}) = p$ ($\hat{p}$ is an *unbiased* estimator of p), and $Var(\hat{p}) = \frac{p(1-p)}{n}$.
2. The normal approximation is best when n is large and $p = 0.5$. The normal approximation is worst when n is small and p is close to 0 or 1.

11.3 Confidence Intervals and Hypothesis Tests for the Population Proportion p

3. Never. We make confidence intervals for *parameters*, and never for *statistics*.
4. (a) $\hat{p} = \frac{62}{200} = 0.31$.

   (b) $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.31(1-0.31)}{200}} = 0.0327$.

   (c)

$$\hat{p} \pm z_{\alpha/2} SE(\hat{p})$$
$$0.31 \pm 1.96 \times 0.0327$$

which works out to 0.31 ± 0.064 or (0.245, 0.374).

(d) $SE_0(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{200}} = 0.03536$.

(e) $Z = \frac{\hat{p}-p_0}{SE_0(\hat{p})} = \frac{0.31-0.5}{0.03536} = -5.374$.

5. (a) $n\hat{p} = 10 \times 0.70 = 7$, $n(1-\hat{p}) = 10 \times 0.30 = 3$. Since these quantities are not both at least 15, we should not be using the normal approximation. (The sampling distribution of $\hat{p}$ cannot be reasonably approximated by a normal distribution in this situation.)

(b) $n\hat{p} = 200 \times 0.70 = 140$, $n(1-\hat{p}) = 200 \times 0.30 = 60$. Since these quantities are both at least 15, we can use the normal approximation. (The sampling distribution of $\hat{p}$ can be reasonably approximated by a normal distribution in this situation.)

(c) $n\hat{p} = 200 \times 0.99 = 198$, $n(1-\hat{p}) = 200 \times 0.01 = 2$. Since these quantities are not both at least 15, we should not be using the normal approximation. (The sampling distribution of $\hat{p}$ cannot be reasonably approximated by a normal distribution in this situation.)

(d) $n\hat{p} = 10{,}000 \times 0.01 = 100$, $n(1-\hat{p}) = 10{,}000 \times 0.99 = 9{,}900$. Since these quantities are both at least 15, we can use the normal approximation. (The sampling distribution of $\hat{p}$ can be reasonably approximated by a normal distribution in this situation.)

(e) Here $n\hat{p} = 0$, and we should not even think about using the normal approximation in this situation.

6. (a) p represents the true proportion of male births to heavy-smoking parents in Liverpool, UK (in the time frame of the study).

(b) $\hat{p} = \frac{158}{363} = 0.435$.

(c) Yes, the sample size is large enough. (There are 158 successes (males) and 205 failures (females), so the *at least 15* guideline is satisfied.)

(d) $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.435(1-0.435)}{363}} = 0.0260$.

(e)

$$\hat{p} \pm 1.96 SE(\hat{p})$$
$$0.435 \pm 1.96 \times 0.0260$$
$$0.435 \pm 0.051$$

which works out to (0.384, 0.486).

(f) We can be 95% confident that the true proportion of male births for Liverpool couples in which both parents are heavy smokers lies between 0.384 and 0.486.

(g) H_0: $p = 0.5$ (The true proportion of male births is 0.50.)
H_a: $p \neq 0.5$ (The true proportion of male births is not 0.50.)
$SE_0(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{363}} = 0.0262$.

$Z = \frac{\hat{p}-p_0}{SE_0(\hat{p})} = \frac{0.435-0.5}{0.0262} = -2.48.$
Since the alternative hypothesis is two-sided, the p-value is double the area to the left of -2.48 under the standard normal curve. The area to the left of -2.48 is 0.0066 (found using software or the standard normal table), and so the p-value is about 0.013.
The p-value is less than the given significance level ($0.013 < 0.05$), so the evidence against the null hypothesis is significant at $\alpha = 0.05$. There is statistically significant evidence that the true proportion of male births (for babies born to heavy-smoking Liverpudlian parents) differs from 0.5. Note that the sample proportion of males was 0.435, with a corresponding 95% confidence interval of (0.384, 0.486), indicating that there is strong evidence that the true proportion of male births is in fact *less* than 0.5.

(h) The sample was of births to Liverpool couples in which both parents were heavy smokers. As such, our conclusions apply only to the population of Liverpool births in which both parents are heavy smokers. (Technically speaking our conclusions apply only to the population from which we sampled, but we may use a study like this to give us a hint of what might be the reality in other scenarios (all UK births to heavy-smoking parents, for example).)

11.4 Determining the Minimum Sample Size n

7. (a) $n = (\frac{1.96}{0.03})^2 \cdot 0.5(1-0.5) = 1067.1$, and we round up to 1068. 1068 is the minimums sample size required.

(b) $n = (\frac{1.96}{0.03})^2 \cdot 0.20(1-0.20) = 682.95$, and we round up to 683. 683 is the minimums sample size required. Note that this is just an approximation. If we are wrong about p being close to 0.2, and it is closer to 0.5, then we will be underestimating the sample size required.

(c) $n = (\frac{1.645}{0.01})^2 \cdot 0.5(1-0.5) = 6765.1$, and we round up to 6,766. 6,766 is the minimums sample size required.

(d) $n = (\frac{2.576}{0.01})^2 \cdot 0.5(1-0.5) = 16589.4$, and we round up to 16,590. 16,590 is the minimums sample size required.

(e) The variance of $\hat{p}$ is greater if $p = 0.2$ than if $p = 0.1$, and thus we would need a larger sample size if $p = 0.2$ than if $p = 0.1$. So, $n = (\frac{2.576}{0.01})^2 \cdot 0.2(1-0.2) = 10617.2$, and we round up to 10,618. 10,618 is the minimums sample size required. (Had we used $p = 0.1$ in the formula instead, we would have arrived at a minimum sample size of 5,973, but this would not have been large enough if p is actually closer to 0.2).

11.5 Inference Procedures for the Difference Between Two Population Proportions

11.5.1 The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

11.5.2 Confidence Intervals and Hypothesis Tests for $p_1 - p_2$

8. $SE(\hat{p}_1 - \hat{p}_2)$ is the standard error of the difference in sample proportions, which is an estimate of the standard deviation of the sampling distribution of $\hat{p}_1 - \hat{p}_2$. It is a measure of the variability in $\hat{p}_1 - \hat{p}_2$ if samples were to be repeatedly drawn from the populations.

9. (a) $\hat{p}_1 = \frac{82}{400} = 0.205$, $\hat{p}_2 = \frac{104}{400} = 0.260$.
   (b) $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0.205(1-0.205)}{400} + \frac{0.260(1-0.260)}{400}} = 0.02981$.
   (c) $0.205 - 0.260 \pm 1.96 \times 0.02981$, which works out to -0.055 ± 0.0584 or $(-0.113, 0.003)$.
   (d) H_0: $p_1 = p_2$, H_a: $p_1 \neq p_2$. The pooled proportion $\hat{p} = \frac{82+104}{400+400} = 0.2325$.
   $SE_0(\hat{p}_1 - \hat{p}_2) = \sqrt{0.2325(1-0.2325)(\frac{1}{400} + \frac{1}{400})} = 0.02987$.
   $Z = \frac{0.205-0.260}{0.02987} = -1.841$. The resulting p-value is 0.066, which does not give significant evidence against the null hypothesis at $\alpha = 0.05$.

10. (a) $\hat{p}_M = \frac{21}{39} = 0.5385$, $\hat{p}_C = \frac{22}{38} = 0.5789$.

$$\begin{aligned} SE(\hat{p}_M - \hat{p}_C) &= \sqrt{\frac{\hat{p}_M(1-\hat{p}_M)}{n_M} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}} \\ &= \sqrt{\frac{0.5385(1-0.5385)}{39} + \frac{0.5789(1-0.5789)}{38}} \\ &= 0.1131 \end{aligned}$$

   The 95% confidence interval for $p_M - p_C$ is:

$$\begin{gathered} \hat{p}_M - \hat{p}_C \pm 1.96 \times SE(\hat{p}_M - \hat{p}_C) \\ 0.5385 - 0.5789 \pm 1.96 \times 0.1131 \\ -0.0404 \pm 0.2216 \end{gathered}$$

   which works out to$(-0.26, 0.18)$.
   (b) We can be 95% confident that the difference in the true proportion of pigeons that will navigate to the home loft under the conditions of the experiment $(p_M - p_C)$ lies between -0.26 and 0.18. (Note that 0 is contained within this interval and thus it is plausible that $p_M = p_C$. The data is consistent with the magnetic pulse having no effect.)

(c) H_0: $p_M = p_C$ (The true proportion of pigeons that will navigate to the home loft under the conditions of this experiment is the same for the magnetic pulse group and the control group. In short, the magnetic pulse has no effect on the likelihood of a pigeon arriving home.)
H_a: $p_M \neq p_C$ (The true proportion of pigeons that will navigate to the home loft under the conditions of this experiment differs between the magnetic pulse group and the control group. In short, the magnetic pulse has an effect on the likelihood of a pigeon arriving home.)

The pooled proportion is $\hat{p} = \frac{21+22}{39+38} = \frac{43}{77} = 0.5584$.

$$\begin{aligned} SE_0(\hat{p}_M - \hat{p}_C) &= \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_M} + \frac{1}{n_C})} \\ &= \sqrt{0.5584(1-0.5584)(\frac{1}{39} + \frac{1}{38})} \\ &= 0.1132 \end{aligned}$$

$$\begin{aligned} Z &= \frac{\hat{p}_M - \hat{p}_C}{SE_0(\hat{p}_M - \hat{p}_C)} \\ &= \frac{0.5385 - 0.5789}{0.1132} \\ &= -0.357 \end{aligned}$$

Since the alternative hypothesis is two-sided, the p-value is double the area to the left of -0.357 under the standard normal curve. The area to the left of -0.357 under the standard normal curve is approximately 0.360, resulting in a p-value of approximately 0.72. This is a large p-value, and it yields absolutely no evidence against the null hypothesis. There is no evidence that the proportion of pigeons that successfully navigate home differs between the magnetic pulse group and the control group. (There is no evidence that the magnetic pulse has any effect on the likelihood of a pigeon navigating back to the loft.)

11.6 Chapter Exercises

11.6.1 Basic Calculations

11. (a) 0.919 (the area to the right of -1.40 under the standard normal curve).
(b) 0.081 (the area to the left of -1.40 under the standard normal curve).
(c) 0.162 (twice the area to the left of -1.40 under the standard normal curve).

(d) 0.030 (the area to the left of -1.88 under the standard normal curve).
(e) 0.060 (twice the area to the left of -1.88 under the standard normal curve).

11.6.2 Concepts

12. $\hat{p}$ is the *sample* proportion, whereas p represents the true proportion for the entire *population*. Once the sample is drawn, we will know the value of $\hat{p}$, but p is typically an unknown value that we are trying to estimate.

13. No, this would never make sense. We test hypotheses about *parameters*, and not about *statistics*.

14. $SE(\hat{p})$ is the standard error of $\hat{p}$, the estimate of the standard deviation of the sampling distribution of $\hat{p}$. It is a measure of the variability in $\hat{p}$, if we were to repeatedly draw samples from the population. Since we regard p as a fixed, unknown quantity that we are trying to estimate, it is simply a constant and it does not have a standard error.

15. (a) $np_0 = 10 \times 0.20 = 2$, $n(1-p_0) = 10 \times 0.80 = 8$. Since these quantities are not both at least 15, we should not be using the normal approximation. (The sampling distribution of $\hat{p}$ cannot be reasonably approximated by a normal distribution in this situation.)
    (b) $np_0 = 200 \times 0.20 = 40$, $n(1-p_0) = 200 \times 0.80 = 160$. Since these quantities are both at least 15, we can use the normal approximation. (The sampling distribution of $\hat{p}$ can be reasonably approximated by a normal distribution in this situation.)
    (c) $np_0 = 10 \times 0.998 = 9.98$, $n(1-p_0) = 10 \times 0.002 = 0.02$. Since these quantities are not both at least 15, we should not be using the normal approximation. (The sampling distribution of $\hat{p}$ cannot be reasonably approximated by a normal distribution in this situation.)
    (d) $np_0 = 1000 \times 0.998 = 998$, $n(1-p_0) = 1000 \times 0.002 = 2$. Since these quantities are not both at least 15, we should not be using the normal approximation. (The sampling distribution of $\hat{p}$ cannot be reasonably approximated by a normal distribution in this situation.)
    (e) $np_0 = 100{,}000 \times 0.998 = 99{,}800$, $n(1-p_0) = 100{,}000 \times 0.002 = 200$. Since these quantities are both at least 15, we can use the normal approximation. (The sampling distribution of $\hat{p}$ can be reasonably approximated by a normal distribution in this situation.)

16. There are many possible sources of bias, including:

    - These types of poll are often conducted by telephone, and will therefore not include anyone who does not have a phone. In the early days of cell phones, these types of surveys called only landline phones. People who were cell-phone-only would be left out (and these people tended to be a younger subgroup of the population, whose opinions may differ in an important way from the population as a whole).

- If this was an online survey, there are obvious biases. People who never use a computer (who tend to be older and poorer than those who do), would not be included.

- The survey does not include people who hang up (refuse to answer questions). If the opinions of these people tend to differ from those who do answer (which is quite likely), then there will be bias in the survey.

- On phone surveys they typically speak to adults only, and thus the interpretation of the survey must relate to adult Canadians, and not Canadians as a whole.

- The people spoken to may have a tendency to lie. As a possible example, a person may feel that schools are as safe as they used to be, but think that if a high proportion say otherwise, perhaps the government will spend more on schools. This type of bias is more obvious with potentially embarrassing questions such as "do you have an STD?" But overall, what people say they think isn't necessarily what they truly think.

17. (a) $\hat{p}$ is an unbiased estimator of p. *True, since* $E(\hat{p}) = p$.
    (b) When $n < 30$ we should use the t distribution when calculating confidence intervals for p. *False. The t distribution never arises in inference for proportions.*
    (c) The true distribution of $\hat{p}$ is based on the binomial distribution. *True.* $\hat{p} = \frac{X}{n}$, *where X has a binomial distribution with parameters n and p.*
    (d) The true standard deviation of the sampling distribution of $\hat{p}$ depends on the value of p. *True.* $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
    (e) The sampling distribution of $\hat{p}$ is perfectly normal for large sample sizes. *False. It's an approximation, but it can be a very good approximation.*

18. (a) The normal approximation to the sampling distribution of $\hat{p}$ works best when we have a large sample size and $p = 0.5$. *True.*
    (b) The sampling distribution of $\hat{p}$ becomes more normal as p tends to 1. *False. The sampling distribution becomes strongly skewed as* $p \to 1$.
    (c) The sampling distribution of p is approximately normal for large sample sizes. *False. p is a parameter, not a statistic, and as such does not have a sampling distribution.*
    (d) All else being equal, the value of $SE(\hat{p})$ decreases as the sample size increases.

True.

(e) In repeated sampling, exactly 95% of 95% confidence intervals for p will capture p. *False, since we are using a normal approximation and not an exact procedure. (But if the sample size is very large then the true percentage will be very close to 95%.)*

11.6.3 Applications

19. (a) p represents the true proportion of homing pigeons that would navigate back to the home loft (for this type of pigeon, under the conditions of this experiment).

(b) $\hat{p} = \frac{22}{38} = 0.5789$.

(c) Yes, the sample size is large enough. There are 22 successes (pigeons that made it back) and 16 failures (pigeons that did not make it back), so the *at least 15* guideline is satisfied (barely).

(d) $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.5789(1-0.5789)}{38}} = 0.0801$.

(e)

$$\begin{aligned}\hat{p} \pm 1.96SE(\hat{p})\\ 0.5789 \pm 1.96 \times 0.0801\\ 0.5789 \pm 0.1570\end{aligned}$$

which works out to (0.42, 0.74).

(f) We can be 95% confident that the true proportion of pigeons that would successfully navigate home lies between 0.42 and 0.74 (for this type of pigeon, under the conditions of this experiment).

(g) H_0: $p = 0.25$ (The true proportion of pigeons that would navigate home is 0.25.)

H_a: $p > 0.25$ (The true proportion of pigeons that would navigate home is greater than 0.25. If we are trying to show strong evidence against the scientist's claim that *no more than 25%* would navigate home on average, we care only about a difference in this direction.)

$SE_0(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{38}} = 0.0702$.

$Z = \frac{\hat{p}-p_0}{SE_0(\hat{p})} = \frac{0.5789-0.25}{0.0702} = 4.68$.

Since the alternative hypothesis is $p > 0.25$, the p-value is the area to the right of 4.68 under the standard normal curve. Using software, we can find: p-value $= 1.4 \times 10^{-6}$. (Using the standard normal table we could say only that the p-value is very close to 0.) This p-value is very small, yielding very strong evidence against the null hypothesis.

There is very strong evidence that, for this type of pigeon under the conditions of this experiment, the true proportion that will successfully navigate to the home loft is greater than 0.25. (This should not come as a great surprise, since we found a 95% confidence interval for p of 0.42 to 0.74.)

20. (a) p represents the true proportion of seatbelt-wearing victims of fatal car accidents that show seat belt marks. (Note that the study involved only accidents in Sydney, Australia, in which there was no airbag deployment. Any conclusions relate to that population of accidents.)
(b) $\hat{p} = \frac{27}{74} = 0.3649$.
(c) Yes, the sample size is large enough. (There are 27 successes (individuals with seat belt marks) and 47 failures (individuals without seat belt marks), so the *at least 15* guideline is satisfied.)
(d) $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.3649(1-0.3649)}{74}} = 0.0560$.
(e)

$$\begin{gathered}\hat{p} \pm 1.96SE(\hat{p}) \\ 0.3649 \pm 1.96 \times 0.0560 \\ 0.3649 \pm 0.1097\end{gathered}$$

which works out to (0.255, 0.475).
(f) We can be 95% confident that p lies between 0.255 and 0.475, where p is the true proportion of seatbelt-wearing victims that show seat belt marks.
(g) No, since the information in the question does not give any indication of a hypothesized value to be tested. It is conceivable that a researcher may have a hypothesis they wish to test, but based on the given information this is an estimation problem, not a testing problem.

21. (a) p represents the true proportion of murders in which the victim was the offender's domestic partner (for Finnish murders in which the murderer was male).
(b) $\hat{p} = \frac{24}{91} = 0.2637$.
(c) Yes, the sample size is large enough. (There are 24 successes (partners) and 67 failures (non-partners), so the *at least 15* guideline is satisfied.)
(d) $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.2637(1-0.2637)}{91}} = 0.0462$.
(e)

$$\begin{gathered}\hat{p} \pm 1.96SE(\hat{p}) \\ 0.2637 \pm 1.96 \times 0.0462 \\ 0.2637 \pm 0.0905\end{gathered}$$

which works out to (0.173, 0.354).
(f) We can be 95% confident that the true proportion of murders that were of the domestic partner of the perpetrator lies between 0.173 and 0.354. (For Finnish murders in which the murderer was male.)

(g) H_0: $p = 0.4$ (The true proportion of victims that were the domestic partner of the perpetrator is 0.40.)
H_a: $p \neq 0.4$ (The true proportion of victims that were the domestic partner of the perpetrator is not 0.40.)
$SE_0(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.4(1-0.4)}{91}} = 0.0514$.
$Z = \frac{\hat{p}-p_0}{SE_0(\hat{p})} = \frac{0.2637-0.4}{0.0514} = -2.65$.
Since the alternative hypothesis is two-sided, the p-value is double the area to the left of -2.65 under the standard normal curve. The area to the left of -2.65 is 0.0040 (found using software or a standard normal table), and so the p-value is about 0.008.
The p-value is less than the given significance level ($0.008 < 0.05$), so the evidence against the null hypothesis is significant at $\alpha = 0.05$. There is statistically significant evidence that the true proportion of victims that were the domestic partner of the perpetrator differs from 0.4. The sample proportion is 0.26, with a 95% confidence interval for p of (0.173, 0.354), indicating strong evidence that the true proportion of victims that are domestic partners is in fact *less* than 0.4.

(h) The sample was of Finnish murder victims in which the perpetrator was male. As such, our conclusions apply only to the population of Finnish murder victims in which the perpetrator was male. (Technically speaking our conclusions apply only to the population from which we sampled, but we may use a study like this to give us a hint of what might be the reality in other scenarios (murders committed by European males, for example).)

22. (a) $\hat{p}_F = \frac{32}{91} = 0.3516$, $\hat{p}_M = \frac{24}{91} = 0.2637$.

$$\begin{aligned} SE(\hat{p}_F - \hat{p}_M) &= \sqrt{\frac{\hat{p}_F(1-\hat{p}_F)}{n_F} + \frac{\hat{p}_M(1-\hat{p}_M)}{n_M}} \\ &= \sqrt{\frac{0.3516(1-0.3516)}{91} + \frac{0.2637(1-0.2637)}{91}} \\ &= 0.0681 \end{aligned}$$

The 90% confidence interval for $p_F - p_M$ is:

$$\hat{p}_F - \hat{p}_M \pm 1.645 \times SE(\hat{p}_F - \hat{p}_m)$$
$$0.3516 - 0.2637 \pm 1.645 \times 0.0681$$
$$0.0879 \pm 0.112$$

which works out to $(-0.024, 0.200)$.

(b) We can be 90% confident that the difference in the true proportions of victims that were domestic partners ($p_F - p_M$) lies between -0.024 and 0.200.

(c) H_0: $p_F = p_M$ (The true proportion of murder victims that were domestic partners is the same for female and male murderers.)

H_a: $p_F \neq p_M$ (The true proportion of murder victims that were domestic partners differs between female and male murderers.)

The pooled proportion is $\hat{p} = \frac{32+24}{91+91} = \frac{56}{182} = 0.3077$.

$$\begin{aligned} SE_0(\hat{p}_F - \hat{p}_M) &= \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_F} + \frac{1}{n_M})} \\ &= \sqrt{0.3077(1-0.3077)(\frac{1}{91} + \frac{1}{91})} \\ &= 0.0684 \end{aligned}$$

$$\begin{aligned} Z &= \frac{\hat{p}_F - \hat{p}_M}{SE_0(\hat{p}_F - \hat{p}_M)} \\ &= \frac{0.3516 - 0.2637}{0.0684} \\ &= 1.285 \end{aligned}$$

Since the alternative hypothesis is two-sided, the p-value is double the area to the right of 1.285 under the standard normal curve. The area to the right of 1.285 under the standard normal curve is about 0.099, resulting in a p-value of approximately 0.20. The p-value is greater than the given significance level ($0.20 > 0.05$), and so the evidence against the null hypothesis is not significant at the 5% significance level. There is not significant evidence that the proportion of murder victims that were domestic partners differs between male and female murderers. (The data is consistent with the two proportions being equal.)

23. (a) $\hat{p}_U = \frac{43}{60} = 0.7167$, $\hat{p}_C = \frac{27}{44} = 0.6136$.

$$\begin{aligned} SE(\hat{p}_U - \hat{p}_C) &= \sqrt{\frac{\hat{p}_U(1-\hat{p}_U)}{n_U} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}} \\ &= \sqrt{\frac{0.7167(1-0.7167)}{60} + \frac{0.6136(1-0.6136)}{44}} \\ &= 0.0937 \end{aligned}$$

The 95% confidence interval for $p_U - p_C$ is:

$$\begin{gathered} \hat{p}_U - \hat{p}_C \pm 1.96 \times SE(\hat{p}_U - \hat{p}_C) \\ 0.7167 - 0.6136 \pm 1.96 \times 0.0937 \\ 0.1031 \pm 0.1835 \end{gathered}$$

which works out to $(-0.08, 0.29)$.

(b) We can be 95% confident that the difference in the true proportions of individuals that express the Le^b antigen ($p_U - p_C$) lies between -0.08 and 0.29. (Note that 0 is contained within this interval and thus it is plausible that $p_U = p_C$. The data is consistent with there being no association between peptic ulcers and the Le^b antigen.)

(c) H_0: $p_U = p_C$ (The true proportion of individuals expressing the Le^b antigen is the same for peptic ulcer sufferers and healthy individuals.)
H_a: $p_U \neq p_C$ (The true proportion of individuals expressing the Le^b antigen differs between peptic ulcer sufferers and healthy individuals.)

The pooled proportion is $\hat{p} = \frac{43+27}{60+44} = \frac{70}{104} = 0.6731$.

$$\begin{aligned} SE_0(\hat{p}_U - \hat{p}_C) &= \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_U} + \frac{1}{n_C})} \\ &= \sqrt{0.6731(1-0.6731)(\frac{1}{60} + \frac{1}{44})} \\ &= 0.0931 \end{aligned}$$

$$\begin{aligned} Z &= \frac{\hat{p}_U - \hat{p}_C}{SE_0(\hat{p}_U - \hat{p}_C)} \\ &= \frac{0.7167 - 0.6136}{0.0931} \\ &= 1.107 \end{aligned}$$

Since the alternative hypothesis is two-sided, the p-value is double the area to the right of 1.107 under the standard normal curve. The area to the right of 1.107 under the standard normal curve is approximately 0.134, resulting in a p-value of approximately 0.27. The p-value is greater than the given significance level ($0.27 > 0.05$), and so the evidence against the null hypothesis is not significant at the 5% significance level. There is not significant evidence that the proportion of individuals expressing the Le^b antigen differs between peptic ulcer sufferers and healthy individuals. (The data is consistent with the two proportions being equal.)

24. (a) $\hat{p}_H = \frac{158}{363} = 0.4353$, $\hat{p}_N = \frac{2685}{5045} = 0.5322$.

$$\begin{aligned} SE(\hat{p}_H - \hat{p}_N) &= \sqrt{\frac{\hat{p}_H(1-\hat{p}_H)}{n_H} + \frac{\hat{p}_N(1-\hat{p}_N)}{n_N}} \\ &= \sqrt{\frac{0.4353(1-0.4353)}{363} + \frac{0.5322(1-0.5322)}{5045}} \\ &= 0.0270 \end{aligned}$$

The 95% confidence interval for $p_H - p_N$ is:

$$\begin{aligned} \hat{p}_H - \hat{p}_N &\pm 1.96 \times SE(\hat{p}_H - \hat{p}_N) \\ 0.4353 - 0.5322 &\pm 1.96 \times 0.0270 \\ -0.0969 &\pm 0.0528 \end{aligned}$$

which works out to $(-0.15, -0.04)$.

(b) We can be 95% confident that the difference between the true proportion of male births to heavy-smoking parents and the true proportion of male births to non-smoking parents ($p_H - p_N$) lies between -0.15 and -0.04. (Note that the entire interval lies to the left of 0, thus giving some evidence that the true proportion of male births for heavy-smoking parents is less than the true proportion for non-smoking parents.)

(c) H_0: $p_H = p_N$ (The true proportion of male births for heavy-smoking parents is equal to the true proportion of male births for non-smoking parents.)
H_a: $p_H \neq p_N$ (The true proportion of male births for heavy-smoking parents differs from the true proportion of male births for non-smoking parents.)

The pooled proportion is $\hat{p} = \frac{158+2685}{363+5045} = \frac{2843}{5408} = 0.5257$.

$$\begin{aligned} SE_0(\hat{p}_H - \hat{p}_N) &= \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_H} + \frac{1}{n_N})} \\ &= \sqrt{0.5257(1-0.5257)(\frac{1}{363} + \frac{1}{5045})} \\ &= 0.0271 \end{aligned}$$

$$\begin{aligned} Z &= \frac{\hat{p}_H - \hat{p}_N}{SE_0(\hat{p}_H - \hat{p}_N)} \\ &= \frac{0.4353 - 0.5322}{0.0271} \\ &= -3.57 \end{aligned}$$

Since the alternative hypothesis is two-sided, the p-value is double the area to the left of -3.57 under the standard normal curve. The area to the left of -3.57 under the standard normal curve is approximately 0.0002 (found using software, or approximated with the standard normal table), resulting in a p-value of approximately 0.0004.

Since the p-value is less than the given significance level ($0.0004 < 0.05$), there is statistically significant evidence against the null hypothesis. There is strong evidence that the true proportion of male births to heavy-smoking parents differs from the true proportion of male births to non-smoking parents. (By looking at the sample proportions of $\hat{p}_H = 0.4353$ and $\hat{p}_N = 0.5322$, and the resulting 95% confidence interval for $p_H - p_N$ of (-0.15, -0.04), we can state that there is strong evidence that the population proportion of male births to heavy-smoking parents is in fact *less* than the population proportion of male births to non-smoking parents.)

(d) The samples were drawn from births in Liverpool, UK, and as such our conclusions apply only to that area in the time frame of the study. (But the results of this study do give us a hint that the observed relationship may hold in other areas as well.) The confidence interval and the hypothesis test give evidence of an *association* between parental smoking and the sex of the child, but they do not imply a *causal* link. (For the usual reasons – this was an observational study, not an experiment, and so there are a multitude of lurking variables that could be the cause of the observed relationship.)

25. (a) $\hat{p}_F = \frac{46}{200} = 0.230$, $\hat{p}_C = \frac{27}{200} = 0.135$.

$$\begin{aligned} SE(\hat{p}_F - \hat{p}_C) &= \sqrt{\frac{\hat{p}_F(1-\hat{p}_F)}{n_F} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}} \\ &= \sqrt{\frac{0.230(1-0.230)}{200} + \frac{0.135(1-0.135)}{200}} \\ &= 0.0383 \end{aligned}$$

The 95% confidence interval for $p_F - p_C$ is:

$$\hat{p}_F - \hat{p}_C \pm 1.96 \times SE(\hat{p}_F - \hat{p}_C)$$
$$0.230 - 0.135 \pm 1.96 \times 0.0383$$
$$0.095 \pm 0.0751$$

which works out to $(0.02, 0.17)$.

(b) We can be 95% confident that the difference between the true success rate in pleasant aroma areas and the true success rate in areas with no aroma ($p_F - p_C$) lies between 0.02 and 0.17. (Note that the entire interval lies to the right of 0, thus giving evidence that the true success rate in pleasant aroma areas is *greater* than the true success rate in areas with no aroma.) There is evidence that the pleasant aroma areas have a positive effect on the success rate. We will now investigate that a little more formally with a hypothesis test.

(c) H_0: $p_F = p_C$ (The true success rate in pleasant aroma areas is equal to the true success rate in areas of no aroma.)
H_a: $p_F \neq p_C$ (The true success rate in pleasant aroma areas differs from the true success rate in areas of no aroma.)

The pooled proportion is $\hat{p} = \frac{46+27}{200+200} = \frac{73}{400} = 0.1825$.

$$
\begin{aligned}
SE_0(\hat{p}_F - \hat{p}_C) &= \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_F} + \frac{1}{n_C})} \\
&= \sqrt{0.1825(1-0.1825)(\frac{1}{200} + \frac{1}{200})} \\
&= 0.0386
\end{aligned}
$$

$$
\begin{aligned}
Z &= \frac{\hat{p}_F - \hat{p}_C}{SE_0(\hat{p}_F - \hat{p}_C)} \\
&= \frac{0.230 - 0.135}{0.0386} \\
&= 2.46
\end{aligned}
$$

Since the alternative hypothesis is two-sided, the p-value is double the area to the right of 2.46 under the standard normal curve. The area to the right of 2.46 under the standard normal curve is approximately 0.007 (found using software or the standard normal table), resulting in a p-value of approximately 0.014.

Since the p-value is less than the given significance level ($0.014 < 0.05$), there is statistically significant evidence against the null hypothesis. There is strong evidence that the true success rate (of obtaining a phone number) in pleasant aroma areas differs from the true success rate in areas with no aroma. (By looking at the sample proportions of $\hat{p}_F = 0.230$ and $\hat{p}_C = 0.135$, and the resulting 95% confidence interval for $p_F - p_C$ of (0.02, 0.17), we can state that there is strong evidence that the true success rate in pleasant aroma areas is *greater* than the true success rate in areas with no aroma.)

Note that the conclusions to this study apply only to the specific conditions of the experiment (good-looking young males approaching young women walking alone in a shopping mall). If the conditions of the experiment were to change (using males who were not quite as good-looking, for example), then the results might possibly change dramatically.

26. (a) $\hat{p}_p = \frac{160}{200430} = 0.0007982837$, $\hat{p}_v = \frac{86}{199747} = 0.0004305446$. The point estimate of the difference in population proportions $(p_p - p_v)$ is the difference in sample proportions: $\hat{p}_p - \hat{p}_v = 0.0007982837 - 0.0004305446 = 0.0003677391$.

(b) The pooled proportion is $\hat{p} = \frac{160+86}{200430+199747} = 0.000614728$.
$SE_0(\hat{p}_p - \hat{p}_v) = \sqrt{0.000614728(1-0.000614728)(1/200430+1/199747)} = 0.00007836327$

$$Z = \frac{0.0007982837 - 0.0004305446}{0.00007836327} = 4.693$$

(c) The p-value is twice the area to the right of 4.693 under the standard normal curve. This area is small (close to 0). By the computer, p-value $= 2.7 \times 10^{-6}$.

(d) Since the p-value is very small (2.7×10^{-6}), there is very strong evidence of a difference in population proportions. Since the sample proportion in the placebo group is larger than that of the vaccine group, there is strong evidence that those in the vaccine group developed polio less often. And since this was a randomized experiment, we can say the strong conclusion: *This study yields strong evidence that use of this vaccine in children reduces the risk of developing polio.*

(e) $RR = \dfrac{0.0007982837}{0.0004305446} = 1.854$. People without the vaccine were 1.85 times more likely to develop polio.

27. (a) $\hat{p} = \frac{68}{100} = 0.68$. $SE(\hat{p}) = \sqrt{\frac{0.68(1-0.68)}{100}} = 0.046648$. The 95% confidence interval is given by $0.68 \pm 1.96 \times 0.046647$, which works out to 0.68 ± 0.0914 or (0.589, 0.771). We can be 95% confident that the proportion of this type of minnow that would die within 48 hours after exposure to 5 ppm of zinc lies between 0.589 and 0.771.

(b) Yes. The entire interval lies to the right of 0.50, indicating that 5 ppm of zinc kills more than 50% of these minnows under these conditions. It appears as though the LD50 is less than 5 ppm.

(c) H_0: $p = 0.50$ (5 ppm of zinc will kill 50% of minnows within 48 hours)
H_a: $p \neq 0.50$ (The true percentage of minnows killed within 48 hours with 5 ppm of zinc differs from 0.50.)
$SE_0(\hat{p}) = \sqrt{\frac{0.50(1-0.50)}{100}} = 0.05$. $Z = \frac{0.68-0.50}{0.05} = 3.6$. The p-value is double the area to the right of 3.6 under the standard normal curve (p-value $= 0.0003$). Since the p-value is less than the given significance level of $\alpha = 0.05$, there is significant evidence against the null hypothesis. There is significant evidence that, under these conditions, the true proportion of minnows killed by 5 ppm of zinc will differ from 0.50. Since the sample proportion is greater than 0.50 (leading to a test statistic in the right tail of the distribution), there is strong evidence that the true proportion killed is in fact greater than 0.50. (Which shouldn't come as a big surprise, since our confidence interval was found to be (0.589, 0.771) in 27a.)

(d) $n = (\frac{2.576}{0.03})^2 \cdot 0.5(1-0.5) = 1843.3$. We would need a sample size of at least 1844 minnows.

(e) No. The rough guideline tells us that we need at least 15 successes and 15 failures for the normal approximation to be reasonable. Here, depending on how we define success and failure, we have 198 of one and only 2 of the other. The sampling distribution of $\hat{p}$ cannot reasonably be approximated by a normal distribution in this spot.

28. (a) $\hat{p} = \frac{43}{67} = 0.64179$. The 95% confidence interval is $0.64179 \pm 1.96\sqrt{\frac{0.64179(1-0.64179)}{67}}$, which works out to $.64179 \pm 0.1148$, or (0.527, 0.757). We can be 95% confi-

dence that the proportion of babies born at 25 weeks that live for at least a year lies between 0.527 and 0.757.

(b) H_0: $p = 0.5$ (Half of 25 week premature babies die before one year.)
H_a: $p \neq 0.5$ (The true proportion premature babies that die before one year differs from 0.5.)
$SE_0(\hat{p}) = \sqrt{\frac{0.5(1-0.5)}{67}} = 0.06108$. $Z = \frac{0.64179-0.5}{0.06108} = 2.321$. The p-value is twice the area to the right of 2.321 under the standard normal curve, which is approximately 0.020. There is very strong evidence (p-value $= 0.02$) that the true proportion of this type of premature baby that survives to one year differs from 0.5. In fact, since the sample proportion is greater than 0.50 (leading to a Z statistic in the right tail of the distribution), there is strong evidence that the true proportion that survives is in fact *greater* than 0.50.

29. $\hat{p} = \frac{17}{40} = 0.425$. $SE(\hat{p}) = \sqrt{\frac{0.425(1-0.425)}{40}} = 0.07816$. The 95% confidence interval is given by $0.425 \pm 1.96 \times 0.07816$, which works out to 0.425 ± 0.1532 or (0.272, 0.578). We can be 95% confident that proportion of all soft-serve ice cream vendors in this city that have *E. coli* counts that exceed recommended guidelines lies between 0.272 and 0.578.

30. (a) $\hat{p} = \frac{25}{500} = 0.05$. $SE(\hat{p}) = \sqrt{\frac{0.05(1-0.05)}{500}} = 0.009746$. The 90% confidence interval is given by $0.05 \pm 1.645 \times 0.009746$, which works out to 0.05 ± 0.016 or (0.034, 0.066). We can be 95% confident that the proportion of adults (in this age group, area, and time) that have a university degree lies between 0.034 and 0.066.

(b) H_0: $p = 0.23$ (The proportion with a university degree was the same as for all of Canada.)
H_a: $p \neq 0.23$ (The proportion with a university degree differed from the rest of Canada.)
$SE_0(\hat{p}) = \sqrt{\frac{0.23(1-0.77)}{500}} = 0.01882$. $Z = \frac{0.05-0.23}{0.01882} = -9.564$. The p-value is double the area to the left of -9.564 under a standard normal curve, which is very near 0 (using a computer, we can find it to be 1.1×10^{-21}). There is extremely strong evidence that the proportion of people that have a university degree in this age group and area differed from that of Canada as a whole. In fact, since the sample proportion is less than the proportion for all of Canada (leading to a test statistic in the left tail of the distribution), there is a strong indication that the true proportion in this area was in fact *less* than that of Canadians as a whole.

31. $n = (\frac{1.645}{0.04})^2 \cdot 0.5(1-0.5) = 422.8$. We would need a sample size of at least 423 individuals in order to estimate p to within 0.04 with 90% confidence.

32. (a) $\hat{p}_F = \frac{58}{100} = 0.58$, $\hat{p}_M = \frac{43}{100} = 0.43$. $SE(\hat{p}_F - \hat{p}_M) = \sqrt{\frac{0.58(1-0.58)}{100} + \frac{0.43(1-0.43)}{100}} =$

0.0699.
A 90% confidence interval for $p_F - p_M$ is given by $0.58 - 0.43 \pm 1.645 \times 0.0699$, which works out to 0.15 ± 0.115 or (0.035, 0.265).

(b) We can be 90% confident that the true difference between the approval ratings of women and men $(p_F - p_M)$ lies between 0.035 and 0.265.

(c) H_0: $p_F = p_M$ (The approval ratings for men and women are equal.)
H_a: $p_F \neq p_M$ (The approval ratings for men and women are different.)
The pooled proportion is $\hat{p} = \frac{58+43}{100+100} = 0.505$.
$SE_0(\hat{p}_F - \hat{p}_M) = \sqrt{0.505(1-0.505)(1/100+1/100)} = 0.070707$. $Z = \frac{0.58-0.43}{0.070707} = 2.121$.
The p-value is twice the area to the right of 2.121 under the standard normal curve, which works out to 0.034. There is significant evidence against the null hypothesis at the given level of $\alpha = 0.05$. There is significant evidence of a true difference in approval ratings between men and women. It appears as though women are more likely to approve of the way the President is handling his job.

33. (a) $\hat{p}_1 = \frac{176}{200} = 0.88$, $\hat{p}_2 = \frac{184}{200} = 0.92$. $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0.88(1-0.88)}{200} + \frac{0.92(1-0.92)}{200}} = 0.02993$.
A 95% confidence interval for $p_1 - p_2$ is given by $0.88 - 0.92 \pm 1.96 \times 0.02993$, which works out to -0.04 ± 0.0587 or $(-0.099, 0.019)$.
We can be 95% confident that the difference in population proportion of bedbugs killed for heat treatments of 45 degrees and 47 degrees Celsius lies between -0.099 and 0.019.

(b) H_0: $p_1 = p_2$ (The proportion of bedbugs killed is the same for both temperatures),
H_a: $p_1 < p_2$ (The proportion of bedbugs killed is lower at the lower temperature).
$\hat{p} = \frac{176+184}{200+200} = 0.90$.
$SE_0(\hat{p}_1 - \hat{p}_2) = \sqrt{0.90(1-0.90)(1/200+1/200)} = 0.030$. $Z = \frac{0.88-0.92}{0.030} = -1.333$.
p-value = 0.091 (the area to the left of -1.333 under the standard normal curve).
There is not significant evidence (at the given significance level of $\alpha = 0.01$) of a difference in proportion of bedbugs killed between the two temperatures.

Chapter 12

Inference for Variances

(No exercises yet.)

Chapter 13

χ^2 Tests for Count Data

J.B.'s strongly suggested exercises: 2, 3, 4, 6, 7, 8, 10, 11, 14, 15, 16, 17, 19, 21, 22, 23

13.1 Introduction

13.2 χ^2 Tests for One-Way Tables

13.2.1 The χ^2 Test Statistic

1. (a) H_0: $p_1 = p_2 = p_3 = p_4 = 0.25$ (The four groups are all equally likely). H_a: $p_i \neq 0.25$ for at least two of the groups (the four groups are not all equally likely).

(b)

Observed	52	149	99	100
Expected	$.25 \times 400 = 100$	$.25 \times 400 = 100$	$.25 \times 400 = 100$	$.25 \times 400 = 100$

(c) $\chi^2 = \dfrac{(52-100)^2}{100} + \dfrac{(149-100)^2}{100} + \dfrac{(99-100)^2}{100} + \dfrac{(100-100)^2}{100} = 47.06.$

(d) $4 - 1 = 3$.

(e) The p-value is the area to the right of 47.06 under a χ^2 distribution with 3 degrees of freedom. The correct area is 3.374954×10^{-10}. From a table, the best we could say is that the area is very small (p-value < 0.0005, or something along those lines).

(f) Yes and yes. With such a small p-value, the null hypothesis would be rejected at any reasonable level.

2. (a) Yes, the observed percentage of Q1 birthdays is higher for Canadian NHL players than for Canadians as a whole (26.5% vs 24.2%). Yes, the observed percentage of Q4 birthdays is lower for Canadian NHL players than for Canadians as a whole (20.6% vs 23.8%).

(b) $H_0 : p_1 = 0.242, p_2 = 0.260, p_3 = 0.260, p_4 = 0.238$ (The distribution of birth dates for Canadian NHL players is the same as for Canadians as a whole.)
H_a : The given probabilities are not all correct. (The distribution of birth dates for Canadian NHL players is not the same as for Canadians as a whole.)

(c) $510 \times 0.242 = 123.42$. This is a fair bit lower than the observed count of 135.

(d) $4 - 1 = 3$.

(e)

Birth date	Q1	Q2	Q3	Q4	Total
Observed count	135	146	124	105	510
Expected count	$.242 \cdot 510 = 123.42$	$.260 \cdot 510 = 132.60$	$.260 \cdot 510 = 132.60$	$.238 \cdot 510 = 121.38$	510

The test statistic:

$$\chi^2 = \frac{(135 - 123.42)^2}{123.42} + \frac{(146 - 132.60)^2}{132.60} + \frac{(124 - 132.60)^2}{132.60} + \frac{(105 - 121.38)^2}{121.38} = 5.209$$

(f) The p-value is the area to the right of 5.209 under a χ^2 distribution with 3 degrees of freedom, as illustrated in Figure 13.1.

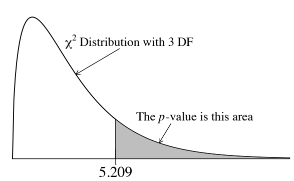


Figure 13.1: The p-value for the NHL birthday question.

Using software, we can find that the p-value is 0.157. Using a χ^2 table, we can find only that the p-value falls in an interval, such as: p-value > 0.10.

(g) Since the p-value is greater than the given significance level ($0.157 > 0.05$), the evidence against H_0 is not significant at $\alpha = 0.05$. There is not significant evidence that the distribution of birth dates for Canadian NHL players differs from that of Canadians as a whole. (It is plausible that the null hypothesis is true.) Note that this does not mean that Malcom Gladwell's assertion is wrong, we simply do not have significant evidence of a birthday effect in this analysis.[1] The observed percentage of Canadian NHL players born in Q1 (26.5) is greater than the percentage for Canadians (24.2). But we do not have enough evidence to say this is a *significant* difference.

[1] A proper attempt to to investigate Canadian NHL player birthdays would involve data over many years, and possibly other methods of analysis. The birthday effect in professional hockey is real, but we simply have not seen much evidence of it here.

13.2.2 Testing Goodness-of-Fit for Specific Parametric Distributions

3. (a) H_0: The number of eggs that hatch in a nest follows a binomial distribution with $n = 3$ and $p = 0.80$. H_a: The number of eggs that hatch in a nest does not follow a binomial distribution with $n = 3$ and $p = 0.80$.
   (b) We find the expected *probabilities* using the binomial distribution with $n = 3$ and $p = 0.80$. The expected probabilities are 0.008, 0.096, 0.384, 0.512. We multiply these by the total number of nests (500) to get the expected counts: 4, 48, 192, 256.
   (c) $\chi^2 = \frac{(80-4)^2}{4} + \frac{(29-48)^2}{48} + \frac{(164-192)^2}{192} + \frac{(227-256)^2}{256} = 1458.889$.
   (d) The degrees of freedom are $4 - 1 = 3$.
   (e) The p-value is the area to the right of 1458.889 under a χ^2 distribution with $4 - 1 = 3$ degrees of freedom. The p-value is close to 0.
   (f) There is very strong evidence that the number of eggs that hatch per nest does not follow a binomial distribution with $p = 0.80$.

4. H_0: The number of eggs that hatch in a nest follows a binomial distribution.
   H_a: The number of eggs that hatch in a nest does not follow a binomial distribution.

   $$\hat{p} = \frac{\text{Number of eggs that hatch}}{\text{Total number of eggs}} = \frac{80\times0+29\times1+164\times2+227\times3}{500\times3} = 0.692.$$

   Using this estimate of p in the binomial formula with $n = 3$, we find the "expected probabilities" for 0, 1, 2, 3 are 0.02921811, 0.19693766, 0.44247034, 0.33137389. Multiplying by the number of nests, we obtain the expected counts: 14.60906, 98.46883, 221.23517, 165.68694. We then use our usual χ^2 statistic:

   $$\chi^2 = \frac{(80-14.60906)^2}{14.60906} + \frac{(29-98.46883)^2}{98.46883} + \frac{(164-221.23517)^2}{221.23517} + \frac{(227-165.68694)^2}{165.68694} = 379.1994.$$

   The p-value is the area to the right of 379.1994 under a χ^2 distribution with $4 - 1 - 1 = 2$ degrees of freedom (recall that we lose one degree of freedom for every parameter that must be estimated from the data – hence the loss of an extra degree of freedom here). This area is near 0.

   There is very, very, very strong evidence that the number of eggs that hatch per nest does not follow a binomial distribution.

5. H_0: $p_1 = p_2 = p_3 = p_4$ (The four possibilities (0 eggs hatch, 1 egg hatches, 2 eggs hatch, 3 eggs hatch) are all equally likely).
   H_a: The four possibilities (0 eggs hatch, 1 egg hatches, 2 eggs hatch, 3 eggs hatch) are not equally likely.

   The expected counts are 125, 125, 125, 125.

$\chi^2 = \frac{(80-125)^2}{125} + \frac{(29-125)^2}{125} + \frac{(164-125)^2}{125} + \frac{(227-125)^2}{125} = 185.328.$

The p-value is the area to the right of 185.328 under a χ^2 distribution with $4-1=3$ degrees of freedom. This area is very close to 0. There is extremely strong evidence that the four possible outcomes are not equally likely.

13.3 χ^2 Tests for Two-Way Tables

6. (a) H_0: There is no association (relationship) between the row and column variables (this can be phrased differently, depending on the sampling design—the null hypothesis may be that the row and column variables are *independent*).

   (b)

	C1	C2
R1	$\frac{200\cdot 650}{1000} = 130$	$\frac{200\cdot 350}{1000} = 70$
R2	$\frac{220\cdot 650}{1000} = 143$	$\frac{220\cdot 350}{1000} = 77$
R3	$\frac{580\cdot 650}{1000} = 377$	$\frac{580\cdot 350}{1000} = 203$

   (c) $\chi^2 = 209.5835$.
   (d) DF $= (3-1)(2-1) = 2$.
   (e) The p-value is the area to the right of 209.5835 under a χ^2 distribution with 2 degrees of freedom. This area is near 0 (p-value $<$ 2.2e-16, as reported by R).
   (f) Yes and yes. With such a tiny p-value, the null hypothesis would be rejected at any reasonable level. There is extremely strong evidence against the null hypothesis.

7. (a) 76.4% of the buyback guns were of small calibre. 20.3% of the homicide guns were of small calibre. (Note that the percentage of small calibre guns is much higher for buyback guns than for homicide guns.)
   (b) H_0: The distribution of gun calibre is the same for buyback guns, homicide guns, and suicide guns.
   H_a: The distribution of gun calibre is not the same for buyback guns, homicide guns, and suicide guns. (In other words, it is not true that the 3 distributions are the same.)
   (c) $\frac{\text{Row total} \times \text{Column total}}{\text{Overall total}} = \frac{834 \times 941}{1435} = 546.89.$
   (d) (Number of rows -1)(Number of columns -1) $= (4-1)(3-1) = 6$.
   (e) The p-value is the area to the right of 422.48 under a χ^2 distribution with 6 degrees of freedom. The p-value is minuscule (very, very close to 0). (The statistical software R reports p-value $< 2.2 \times 10^{-16}$, which is R's way of saying that the p-value is very, very close to 0.)
   (f) Here the p-value is minuscule, indicating extremely strong evidence against the

null hypothesis. (It would be nearly impossible to observe what was observed, if the null hypothesis were true.) There is extremely strong evidence that the distribution of gun calibre differs between the 3 types of gun (buybacks, homicides, suicides).

(g) It appears as though buyback guns were more likely than homicide and suicide guns to be of smaller calibre, and homicide and suicide guns were more likely than buyback guns to be of medium or large calibre. (We could investigate this in greater detail using statistical inference techniques.)

(h) The samples were of guns from buyback programs in Milwaukee, and guns used in homicides and suicides in Milwaukee, so those are the populations to which the conclusions apply.

13.4 Chapter Exercises

13.4.1 Basic Calculations

13.4.2 Concepts

8. E. The χ^2 approximation breaks down when one or more of the expected counts is small. Depending on how small they are, and how many are small, the approximation may be very poor. (Some people use the rough guideline that the approximation is poor if some of the expected counts are less than 5. But this guideline is quite conservative—if one or two of the expected counts slip a little under 5, it is not a big problem.)

9. No. Here we have 50 children, but the sum of the observed is 119. The children are appearing in more than one cell. (In other words, these 119 observations are not independent.) The χ^2 test is not appropriate here.

10. (a) 0.005.
    (b) 0.15.
    (c) $P(Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 > 7.779) = 0.10$. Here we are summing four squared independent standard normal random variables. This sum has a χ^2 distribution with 4 degrees of freedom (we are summing 4 squared terms). $P(Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 > 7.779) = 0.10$ (found from the χ^2 distribution with 4 degrees of freedom, using either software or the χ^2 table).
    (d) $P(Z_1^2 + Z_2^2 + Z_3^2 < 11.345) = 0.99$. Here we are summing three squared independent standard normal random variables. This sum has a χ^2 distribution with 3 degrees of freedom (we are summing 3 squared terms). $P(Z_1^2+Z_2^2+Z_3^2 > 11.345) = 0.01$, $P(Z_1^2 + Z_2^2 + Z_3^2 < 11.345) = 0.99$ (found from the χ^2 distribution with 3 degrees of freedom, using either software or the χ^2 table).

11. (a) True. The p-value is the area to the *right* of the test statistic.
    (b) False. As the degrees of freedom increase, the χ^2 distribution gets closer to a *normal* distribution, that part is true. But the mean of a χ^2 distribution is equal to its degrees of freedom (not 0, as for the standard normal distribution).
    (c) True. (But the skewness decreases as the degrees of freedom increase.)
    (d) True.
    (e) False. (Remember that the χ^2 test statistic is the sum of squared terms.)

12. (a) False, it is only an approximation.
    (b) True.
    (c) True.
    (d) True.
    (e) True.
    (f) True.

13.4.3 Applications

13. (a) H_0: True ratio of phenotypes is 9:3:3:1
    H_a: True ratio of phenotypes is not 9:3:3:1

(b)

	Phenotype			
	Tall/cut	Tall/potato	Dwarf/cut	Dwarf/potato
Observed count	926	288	293	104
Expected under 9:3:3:1	$\frac{9}{16} \times 1611 = 906.2$	$\frac{3}{16} \times 1611 = 302.1$	$\frac{3}{16} \times 1611 = 302.1$	$\frac{1}{16} \times 1611 = 100.7$

Table 13.1: Observed and expected counts of phenotypes in the second generation.

(The given values are rounded, but you should carry many decimal places throughout your calculations.)

$$\chi^2 = \frac{(926-906.2)^2}{906.2} + \frac{(288-302.1)^2}{302.1} + \frac{(293-302.1)^2}{302.1} + \frac{(104-100.7)^2}{100.7} = 1.469$$

    (c) 0.69. The p-value is the area to the right of 1.469 under the χ^2 distribution with $4-1=3$ degrees of freedom. Using software, we can find that the p-value is 0.69. (Using a χ^2 table, we could find only a range for the p-value, such as p-value > 0.20.)
    (d) Since the p-value is large (0.69), there is no evidence against the null hypothesis. The null hypothesis would not be rejected at any reasonable significance level. There is no evidence that the true ratio of phenotypes differs from 9:3:3:1. This experiment gives no evidence of genetic linkage between the genes.

14. (a) H_0: $p = 0.75$ (The true proportion of round seeds is 0.75.)

H_a: $p \neq 0.75$ (The true proportion of round seeds is not 0.75.)

(b) Of the 80 plants, 72 had round seeds and 8 had wrinkled seeds (these are the *observed counts*). If the null hypothesis is true, we would expect to get, on average, $0.75 \times 80 = 60$ plants with round seeds, and $0.25 \times 80 = 20$ with wrinkled seeds.
$\chi^2 = \frac{(72-60)^2}{60} + \frac{(8-20)^2}{20} = 9.6$.

(c) The p-value is the area to the right of 9.6 under the χ^2 distribution with $2 - 1 = 1$ degrees of freedom. Using software, the p-value can be found to be 0.0019. (Using a χ^2 table, we could give only a range, such as $0.001 <$ p-value < 0.005.)

(d) The p-value is less than the given significance level ($0.0019 < 0.05$) and so the evidence against the null hypothesis is statistically significant at $\alpha = 0.05$. There is very strong evidence that the true proportion of plants with round seeds differs from 0.75. (Note that the observed proportion of round seeds ($\frac{72}{80} = 0.90$) is greater than the hypothesized proportion. It appears as though the true proportion of round seeds is greater than 0.75.)

15. (a) $\hat{p} = \frac{72}{80} = 0.90$. $Z = \frac{0.90-0.75}{\sqrt{\frac{.75(1-.75)}{80}}} = 3.098$.

(b) p-value $= 0.0019$. (The p-value is double the area to the right of 3.098 under the standard normal curve, which can be found using software or the standard normal table.)

(c) The square of the Z test statistic is exactly equal to the χ^2 test statistic. ($3.098^2 = 9.6$ (other than a little rounding error).)

(d) The p-values are exactly equal. The two tests are equivalent.

16. (a) H_0: The first digit has a uniform distribution.
H_a: The first digit does not have a uniform distribution.
Alternatively:
H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = \frac{1}{9}$.
H_a: These probabilities are not all correct.

(b) # of categories $- 1 = 9 - 1 = 8$.

(c) p-value $= 5.6 \times 10^{-8}$. The p-value is the area to the right of 49.29 under a χ^2 distribution with 8 degrees of freedom. Using software we can find that this area is 5.6×10^{-8}. (Using a χ^2 table, we could say only that the p-value is very near 0.)

(d) The p-value of 5.6×10^{-8} is tiny, indicating extremely strong evidence against the null hypothesis. There is extremely strong evidence that the first digit does not follow a uniform distribution.

(e) The figure shows that the smaller values (such as 1 and 2) are more likely than the uniform distribution would predict, and that the larger values (such as 7, 8, and 9) are less likely than the uniform distribution would predict.

(f) H_0: The first digit follows Benford's law.

H_a: The first digit does not follow Benford's law.
Alternatively:
H_0: $p_1 = 0.30, p_2 = 0.18, p_3 = 0.12, p_4 = 0.10, p_5 = 0.08, p_6 = 0.07, p_7 = 0.06, p_8 = 0.05, p_9 = 0.05$
H_a: These probabilities are not all correct.

(g) # of categories − 1 = 9 − 1 = 8.

(h) p-value = 0.82. The p-value is the area to the right of 4.41 under a χ^2 distribution with 8 degrees of freedom. Using software we can find that this area is 0.82. (Using a χ^2 table, we could only give a range, such as p-value > 0.20.)

(i) The p-value is large (0.82), indicating no evidence against the null hypothesis. There is no evidence that the distribution of the first digit differs from Benford's law. (The observed data is consistent with Benford's law being the true distribution.)

(j) Yes. The hypothesis test showed no evidence that the distribution differed from Benford's law, and the figure shows that the predicted values from Benford's law are very close to what was observed in the sample.

17. (a) H_0: ABO blood type and Rh factor are independent.
H_a: ABO blood type and Rh factor are not independent.

(b) $\frac{\text{Row total} \times \text{Column total}}{\text{Overall total}} = \frac{39 \times 101}{305} = 12.91$.

(c) (Number of rows − 1)(Number of columns − 1) = $(2-1)(4-1) = 3$.

(d) p-value $= 5.5 \times 10^{-6}$ (found using software). If we used a table instead of software we could only give a range for the p-value, such as p-value < 0.0005.

(e) Since the p-value is less than the given significance level ($5.5 \times 10^{-6} < 0.05$), we can say that the evidence against the null hypothesis is statistically significant at $\alpha = 0.05$. (The p-value is much less than 0.05, so the evidence against the null hypothesis would be significant at much smaller values of α, but we only consider the given α level.) This means there is very strong evidence that ABO blood type and Rh factor are *not* independent (among Libyan medical school students).

(f) The authors of the study report that this was a random sample of Libyan medical school students from a certain university in Libya. Thus (strictly speaking) our conclusions apply only to the population of Libyan medical school students at this university.

(g) This is an unusual result because ABO blood type and Rh factor are controlled by different genes that assort independently. So, at conception at least, ABO and Rh are independent variables. If these are truly independent variables, then it's unusual to see such a large χ^2 test statistic and resulting tiny p-value.

18. (a) 18.3% of the 60 of peptic ulcer patients have blood type A. 34.1% of the 44 healthy individuals have blood type A.

(b) H_0: The distribution of ABO blood type is the same for peptic ulcer sufferers

and healthy individuals.
H_a: The distribution of ABO blood type is different for peptic ulcer sufferers and healthy individuals.

(c) $\frac{\text{Row total} \times \text{Column total}}{\text{Overall total}} = \frac{60 \times 26}{104} = 15$.

(d) (Number of rows -1)(Number of columns -1) $= (2-1)(4-1) = 3$.

(e) The p-value is the area to the right of 7.0521 under a χ^2 distribution with 3 degrees of freedom. p-value $= 0.07026$ (found using software). If we used a table instead of software we could only give a range for the p-value, such as $0.05 < p\text{-value} < 0.10$.

(f) Since the p-value is greater than the given significance level of 0.05, the evidence against the null hypothesis is not statistically significant at this level. We do not have strong evidence of a difference in the distribution of ABO blood type between peptic ulcer sufferers and healthy individuals. (It is plausible that the distribution of ABO blood type is the same for the two groups.) The smallish p-value of 0.07 gives a hint of evidence against the null hypothesis, but it is not statistically significant at $\alpha = 0.05$.

19. (a) The proportion of male births to non-smoking couples is $\frac{2685}{5045} = 0.532$. The proportion of male births for couples where the father is a non-smoker but the mother is a heavy smoker is $\frac{175}{365} = 0.479$.

(b) H_0: Sex of the baby and maternal smoking status are independent variables.
H_a: Sex of the baby and maternal smoking status are not independent variables.

(c) $\frac{\text{Row total} \times \text{Column total}}{\text{Overall total}} = \frac{3595 \times 5045}{6853} = 2646.545$.

(d) (Number of rows -1)(Number of columns -1) $= (2-1)(3-1) = 2$.

(e) The p-value is the area to the right of 5.499 under a χ^2 distribution with 2 degrees of freedom. p-value $= 0.064$ (found using software). If we used a table instead of software we could only give a range for the p-value, such as $0.05 < p\text{-value} < 0.10$.

(f) Since the p-value is greater than the given significance level of 0.05, the evidence against the null hypothesis is not statistically significant at this level. We do not have strong evidence that male births and maternal smoking status are not independent. (It is plausible that they are independent.) The smallish p-value of 0.064 gives a hint of evidence against the null hypothesis, but it is not statistically significant at $\alpha = 0.05$.

(g) The sample involved births to parents in Liverpool, UK, in which the father was a non-smoker. As such, our conclusions apply only to the population of Liverpool births in which the father is a non-smoker. (Technically speaking our conclusions apply only to the population from which we sampled, but we may use a study like this to give us a hint of what might be the reality in other scenarios (all UK births, for example).)

20. (a) The proportion of male births to non-smoking couples is $\frac{2685}{5045} = 0.532$. The proportion of male births for couples where the mother is a non-smoker but the father is a heavy smoker is $\frac{122}{245} = 0.498$.
    (b) H_0: Sex of the baby and paternal smoking status are independent variables.
    H_a: Sex of the baby and paternal smoking status are not independent variables.
    (c) $\frac{\text{Row total}\times\text{Column total}}{\text{Overall total}} = \frac{3192\times5045}{6032} = 2669.7$.
    (d) (Number of rows -1)(Number of columns -1) $= (2-1)(3-1) = 2$.
    (e) The p-value is the area to the right of 1.461 under a χ^2 distribution with 2 degrees of freedom. p-value $= 0.482$ (found using software). If we used a table instead of software we could only give a range for the p-value, such as p-value > 0.10.
    (f) Here the p-value is large (0.48), indicating absolutely no evidence against the null hypothesis. The observed data is consistent with sex of the baby and paternal smoking status being independent variables.
    (g) The sample involved births to parents in Liverpool, UK, in which the mother was a non-smoker. As such, our conclusions apply only to the population of Liverpool births in which the mother is a non-smoker. (Technically speaking our conclusions apply only to the population from which we sampled, but we may use a study like this to give us a hint of what might be the reality in other scenarios (all UK births, for example).)

21. (a) For female murderers, 19.8% of the murders were of family members. For male murderers, 4.4% of the murders were of family members. (Note that the percentage for female murderers is much higher.)
    (b) H_0: The distribution of relationship to the victim is the same for male and female murderers.
    H_a: The distribution of relationship to the victim is different for male and female murderers.
    (c) $\frac{\text{Row total}\times\text{Column total}}{\text{Overall total}} = \frac{22\times91}{182} = 11$.
    (d) (Number of rows -1)(Number of columns -1) $= (4-1)(2-1) = 3$.
    (e) The p-value is the area to the right of 16.638 under a χ^2 distribution with 3 degrees of freedom. p-value $= 0.00084$ (found using software). If we used a table instead of software we could only give a range for the p-value, such as p-value < 0.001.
    (f) Here the p-value is very small (0.00083), indicating very strong evidence against the null hypothesis. There is very strong evidence that the distribution of relationship to the victim for female murderers differs from the distribution of relationship to the victim for male murderers.
    (g) It appears as though females were more likely than males to murder family members or partners, and males were more likely than females to murder acquaintances. (We could investigate this in greater detail using plots and statistical inference techniques.)

(h) The samples were of convicted male and female murderers in Finland, so those are the populations to which the conclusions apply.

22. (a) $H_0 : p_{placebo} = p_{vaccine}$ (the vaccine has no effect).
$H_a : p_{placebo} \neq p_{vaccine}$ (the vaccine has an effect).
$\chi^2 = 22.02188$. The p-value is the area to the right of this value under a χ^2 distribution with 1 degree of freedom. By computer: p-value $= 2.7 \times 10^{-6}$. By the table: p-value < 0.0005.
There is very strong evidence (p-value $= 2.7 \times 10^{-6}$) of a relationship between vaccine use and the rate of polio. From an inspection of the values in the table, we can see that individuals that received the vaccine were *less likely* to have gotten polio. In addition, since this was a well-designed experiment, there is very strong evidence that the polio vaccine *causes* a reduction in the rate of polio.
(b) $4.6927^2 = 22.02188$ (other than a little round-off error). In general for 2×2 tables of count data, $Z^2 = \chi^2$, and the Z and χ^2 tests are equivalent tests.

23. (a) H_0: Text colour and ID are independent variables. (The likelihood of a student filling out their ID number incorrectly is the same for the two colours.)
H_a: Text colour and ID not independent variables. (The likelihood of a student filling out their ID number incorrectly is different between the colours.)
(b) Expected number of white-test writers that incorrectly fill out their ID: $\frac{17\times 423}{842} = 8.54$
Expected number of yellow-test writers that incorrectly fill out their ID: $\frac{17\times 419}{842} = 8.46$
Expected number of white-test writers that properly fill out their ID: $\frac{825\times 423}{842} = 414.46$
Expected number of yellow-test writers that properly fill out their ID: $\frac{825\times 419}{842} = 410.54$

$$\chi^2 = \frac{(1-8.54)^2}{8.54} + \frac{(16-8.46)^2}{8.46} + \frac{(422-414.46)^2}{414.46} + \frac{(403-410.54)^2}{410.54} = 13.654$$

(c) Using software, the p-value is 0.00022. (Using the χ^2 table we could give only a range of values, such as p-value < 0.0005.
(d) The observed data does give strong evidence against the null hypothesis (p-value = 0.00022). For a few reasons, I believe the observed data was simply a fluke and the observed effect is not a real one. Consider the following factors:
- The hypothesis test was only being carried out because the observed data was unusual and unexpected. We should not use the same data that suggests a hypothesis to test that hypothesis, as that can lead to a misleading p-value, and misleading conclusions. (It would have been far more convincing had I planned to test this hypothesis before the students wrote the test.)
- There is no plausible explanation for why yellow-test writers would be

more likely to fill out an incorrect ID number. (It may be the case that the observed effect is a real one, but it would be much more convincing if we had a plausible explanation for the effect that was seen in the sample.)

- This effect was not seen in many other tests, including tests with the same students in the same course.

13.4.4 Extra Practice Questions

24. (a) H_0: $p_1 = p_2 = p_3 = \frac{1}{3}$ (the pictures are equally preferred). H_a: The probabilities are not all equally likely (the pictures are not equally preferred).
    (b) Observed counts: 24, 37, 39. Expected counts: 33.333, 33.333, 33.333.
    $\chi^2 = \frac{(24-33.333)^2}{33.333} + \frac{(37-33.333)^2}{33.333} + \frac{(39-33.333)^2}{33.333} = 3.98$.
    (c) The p-value is the area to the right of 3.98 under a χ^2 distribution with 2 degrees of freedom. Using a computer, p-value $= 0.1367$. Using a table, we could give only a range, such as: $0.10 < p\text{-value} < 0.15$.
    (d) There is very little evidence ($0.10 < p\text{-value} < 0.15$) against the hypothesis that the three pictures are equally preferred.

25. (a) H_0: The true ratio is 1:1:2:2. (3 and 4 stars are twice as likely as 1 and 2 stars).
    H_a: The true ratio is not 1:1:2:2.
    (b) Observed counts: 12, 79, 209, 100. Expected counts: 66.667, 66.667, 133.333, 133.333.
    $\chi^2 = \frac{(12-66.667)^2}{66.667} + \frac{(79-66.667)^2}{66.667} + \frac{(209-133.333)^2}{133.333} + \frac{(100-133.333)^2}{133.333} = 98.38$.
    (c) The p-value is the area to the right of 98.38 under a χ^2 distribution with $4 - 1 = 3$ degrees of freedom. This area is near 0.
    (d) There is extremely strong evidence against the movie executive's claim. The probability of seeing what was seen in this sample, if the executive's claim is true, is near 0.

26. H_0: $p_1 = p_2 = p_3 = \frac{1}{3}$ (the 3 flavours are equally preferred).
    H_a: $p_i \neq \frac{1}{3}$ for at least one i (the 3 flavours are not equally preferred).

    If the 3 flavours are equally preferred, we would expect to get 40 observations in each cell on average.

    $$\chi^2 = \frac{(26-40)^2}{40} + \frac{(67-40)^2}{40} + \frac{(27-40)^2}{40} = 27.35.$$

    The p-value is the area to the right of 27.35 under a χ^2 distribution with $3 - 1 = 2$ degrees of freedom. Using software: p-value $= 1.15 \times 10^{-6}$. Using the table: p-value < 0.0005.

There is very strong evidence (p-value $= 1.15 \times 10^{-6}$, significant at $\alpha = 0.05$) that the flavours are not equally preferred.

27. The p-value is the area to the right of 377.42 under a χ^2 distribution with $(3-1)(4-1) = 6$ degrees of freedom. This area is near 0. There is extremely strong evidence that age and rating are not independent for viewers of this movie.

28. C is the best option. The tiny p-value gives very strong evidence of an association between the type of concrete and the type of damage.

29. The χ^2 test statistic works out to 97.4709.

    The p-value is the area to the right of 97.4709 under a χ^2 distribution with $(2-1) \times (3-1) = 2$ degrees of freedom. Using software, we can find that the p-value is very near 0 (so close to 0 that most software will simply report it as 0). Using the χ^2 table, we can find that the p-value is much less than 0.0005.

    The null hypothesis is that there is no association between the row and column variables (in other words, that the methods are all equally effective at killing bedbugs). Since we have a tiny p-value, there is very strong evidence against the null hypothesis. As such, there is very strong evidence that the methods are not all equally effective at killing bedbugs.

Chapter 14

One-Way ANOVA

J.B.'s strongly suggested exercises: 2, 3, 5, 8, 10, 12, 13, 16, 18, 19, 22, 23, 25, 26, 28

14.1 Introduction

14.2 One-Way ANOVA

1. (a) H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ (The population means of the k groups are all equal.)
   (b) H_a: $\mu_i \neq \mu_j$ for at least one i, j pair. (The population means of the k groups are not all equal.)

14.3 Carrying out the One-Way ANOVA

14.3.1 The Formulas

2. (a)

Source	DF	SS	MS	F	p-value
Groups	$4 - 1 = 3$	800	$\frac{800}{3} \approx 266.667$	$\frac{(800/3)}{(200/36)} = 48$	p-value < 0.001
Error	$40 - 4 = 36$	$1000 - 800 = 200$	$\frac{200}{36} \approx 5.556$	—	—
Total	$40 - 1 = 39$	1000	—	—	—

   (b) $MSE = s_p^2 \approx 5.556$.
   (c) $F = 48$.
   (d) Yes (since p-value < 0.05).

3. (a) $s_p^2 = MSE = \frac{SSE}{n-k}$, where $SSE = \sum(n_i-1)s_i^2 = (10-1)2.1^2 + (10-1)4.2^2 + (8-1)3.7^2 = 294.28$. $MSE = \frac{294.28}{28-3} = 11.7712$.
   (b) $SST = \sum n_i(\bar{X}_i - \bar{X})^2 = 10(3.1-10.11429)^2 + 10(10.1-10.11429)^2 + 8(18.9-10.11429)^2 = 1109.514$.
   (c)

Source	DF	SS	MS	F	p-value
Groups	$3-1=2$	1109.514	$\frac{1109.514}{2} = 554.757$	$\frac{554.757}{11.7712} = 47.12833$	p-value < 0.001
Error	$28-3=25$	294.28	11.7712	—	—
Total	$28-1=27$	1403.794	—	—	—

   (d) p-value < 0.001
   (e) There is very strong evidence that the three groups do not all have the same population mean. (The given significance level is 0.05, and since the p-value is less than 0.05, we can say that the evidence against H_0 is significant at the 5% level.)

4. (a) $\bar{X}_1 = 5$, $\bar{X}_2 = 12$, $\bar{X}_3 = 6$, $n_1 = 2$, $n_2 = 3$, $n_3 = 2$, $s_1^2 = 2$, $s_2^2 = 16$, $s_3^2 = 8$.

   $\bar{X} = \frac{58}{7} = 8.285714$.
   $SST = 2\times(5-8.285714)^2 + 3\times(12-8.285714)^2 + 2\times(6-8.285714)^2 = 73.42857$.
   (b) $$s_p^2 = MSE = \frac{(2-1)2 + (3-1)16 + (2-1)8}{2+3+2-3} = 10.5.$$
   $MST = \frac{SST}{k-1} = \frac{73.42857}{2} = 36.71428$.
   (c) $F = \frac{MST}{MSE} = \frac{36.71428}{10.5} = 3.497$.

14.3.2 An Example with Full Calculations

5. (a) Yes, ANOVA is appropriate here. We have a categorical explanatory variable with 3 levels (truth telling, unrehearsed lying, and rehearsed lying), and a quantitative response variable (eye movement). We are interested in investigating possible differences in the mean response between the groups, so our goals mesh with the goals of ANOVA.

   The standard deviations for the different groups are very similar (36, 44, 38), so the common variance assumption appears to be reasonable. The sample sizes are large, so the normality assumption is not that important. Overall it's very reasonable to use ANOVA here. But we should plot the data, as this can give us an overall view of what we are dealing with, help to assess normality, and show any outliers or other problems.
   (b) It's very reasonable to use ANOVA here, even if there is a little right skewness. ANOVA works well, even if there is a little non-normality in the groups (especially if the sample sizes are large, and they are fairly large here). We should

plot the data and investigate the skewness though—if the skewness is strong we might consider using a transformation before carrying out the ANOVA. (For example, we might consider carrying out an ANOVA on the logs of the eye movement measurements.)

(c) There is statistically significant evidence (p-value $= 0.019 < 0.05$) that the groups do not all have the same population mean eye movement. (In other words, there is statistically significant evidence that the effect on eye movement differs between the treatment groups.)

(d) No, the analysis is not yet complete. Since there is statistically significant evidence that the groups do not all have the same true mean, we will investigate the differences between the groups using pairwise comparisons.

14.4 What Should We Do After a One-Way ANOVA?

14.4.1 Introduction

6. $0.95^{25} = 0.277$.

7. $(1 - 0.05)^{10} = 0.5987$.

14.4.2 Fisher's LSD Method

8. (a) There are $\binom{3}{2} = 3$ pairwise comparisons. We will estimate $\mu_{UL} - \mu_{TT}$, $\mu_{RL} - \mu_{TT}$ and $\mu_{UL} - \mu_{RL}$.

(b) A 95% confidence interval for $\mu_i - \mu_j$ is $\bar{X}_i - \bar{X}_j \pm t_{.025} SE(\bar{X}_i - \bar{X}_j)$, where $SE(\bar{X}_i - \bar{X}_j) = \sqrt{MSE}\sqrt{1/n_i + 1/n_j}$. The t has 122 degrees of freedom (the degrees of freedom for error), yielding $t_{.025} = 1.9796$ (found using software – the value may be slightly different if you look it up in a t table).
The interval for $\mu_{UL} - \mu_{TT}$ is:

$$\begin{aligned}
&\bar{X}_{UL} - \bar{X}_{TT} \pm t_{.025} SE(\bar{X}_{UL} - \bar{X}_{TT}) \\
&\quad 63 - 42 \pm 1.9796 \times \sqrt{1561}\sqrt{1/42 + 1/41} \\
&\quad 21 \pm 17.17
\end{aligned}$$

which works out to $(3.8, 38.2)$.
The interval for $\mu_{RL} - \mu_{TT}$ is:

$$\begin{aligned}
&\bar{X}_{RL} - \bar{X}_{TT} \pm t_{.025} SE(\bar{X}_{RL} - \bar{X}_{TT}) \\
&\quad 64 - 42 \pm 1.9796 \times \sqrt{1561}\sqrt{1/42 + 1/41} \\
&\quad 22 \pm 17.17
\end{aligned}$$

which works out to $(4.8, 39.2)$.
The interval for $\mu_{UL} - \mu_{RL}$ is:

$$\begin{aligned}\bar{X}_{UL} - \bar{X}_{RL} \pm t_{.025} SE(\bar{X}_{RL} - \bar{X}_{TT})\\ 63 - 64 \pm 1.9796 \times \sqrt{1561}\sqrt{1/42 + 1/42}\\ -1 \pm 17.07\end{aligned}$$

which works out to $(-18.1, 16.1)$.

(c) The interval for $\mu_{UL} - \mu_{RL}$ contains 0, and so there is not statistically significant evidence (at $\alpha = 0.05$) of a difference in mean eye movement between these groups.
The intervals for $\mu_{UL} - \mu_{TT}$ and $\mu_{RL} - \mu_{TT}$ do not contain 0, and thus there is statistically significant evidence that μ_{UL} and μ_{RL} both differ from μ_{TT}. Since both intervals lie entirely to the *right* of 0, we have strong evidence that μ_{UL} and μ_{RL} are both *greater* than μ_{TT}

(d) The true family-wise confidence level is unknown, as it depends on the dependency between the comparisons, but it will be less than 95%. For a little perspective, if the comparisons were independent (they are not), the family-wise confidence level would be $0.95^3 \times 100\% = 85.7\%$.

(e) From the ANOVA, there is statistically significant evidence (p-value $= 0.019 < 0.05$) that the different lie conditions do not all result in the same true mean eye movement. 95% LSD confidence intervals showed that both the unrehearsed lying group and the rehearsed lying group had mean eye movements that were significantly greater than that of the truth tellers, but there was not a statistically significant difference between the rehearsed and unrehearsed liars.

14.4.3 The Bonferroni Correction

9. (a) $\binom{5}{2} = 10$.
   (b) If we want to keep α at no more than 0.10, for the individual intervals we use $\alpha' = \frac{0.10}{10} = 0.01$. Each interval should have a 99% confidence level.
   (c) 2.787. $t_{0.01/2} = t_{0.005}$. There are $n - k = 30 - 5 = 25$ degrees of freedom. $t_{0.005} = 2.787$.

10. (a) There are $\binom{3}{2} = 3$ pairwise comparisons. We will estimate $\mu_{UL} - \mu_{TT}$, $\mu_{RL} - \mu_{TT}$ and $\mu_{UL} - \mu_{RL}$.
   (b) The family-wise confidence level of 95% implies $\alpha = 0.05$. For the 3 individual intervals, $\alpha' = \frac{0.05}{3}$, with a resulting confidence interval of $(1 - \frac{0.05}{3})100\% = 98.33\%$.
   (c) The appropriate t value is $t_{\alpha'/2} = t_{.05/6} = t_{0.00833}$. With 122 degrees of freedom (the degrees of freedom for error), $t_{0.00833} = 2.4274$ (found using software – it

cannot be found from the t table). A 95% confidence interval for $\mu_i - \mu_j$ is $\bar{X}_i - \bar{X}_j \pm t_{0.00833} SE(\bar{X}_i - \bar{X}_j)$, where $SE(\bar{X}_i - \bar{X}_j) = \sqrt{MSE}\sqrt{1/n_i + 1/n_j}$. The interval for $\mu_{UL} - \mu_{TT}$ is:

$$\begin{aligned}\bar{X}_{UL} - \bar{X}_{TT} &\pm t_{0.00833} SE(\bar{X}_{UL} - \bar{X}_{TT})\\ 63 - 42 &\pm 2.4274 \times \sqrt{1561}\sqrt{1/42 + 1/41}\\ 21 &\pm 21.06\end{aligned}$$

which works out to $(-0.06, 42.06)$.
The interval for $\mu_{RL} - \mu_{TT}$ is:

$$\begin{aligned}\bar{X}_{RL} - \bar{X}_{TT} &\pm t_{0.00833} SE(\bar{X}_{RL} - \bar{X}_{TT})\\ 64 - 42 &\pm 2.4274 \times \sqrt{1561}\sqrt{1/42 + 1/41}\\ 22 &\pm 21.06\end{aligned}$$

which works out to $(0.94, 43.06)$.
The interval for $\mu_{UL} - \mu_{RL}$ is:

$$\begin{aligned}\bar{X}_{UL} - \bar{X}_{RL} &\pm t_{0.00833} SE(\bar{X}_{RL} - \bar{X}_{TT})\\ 63 - 64 &\pm 2.4274 \times \sqrt{1561}\sqrt{1/42 + 1/42}\\ -1 &\pm 20.93\end{aligned}$$

which works out to $(-21.93, 19.93)$.

(d) The 95% family-wise intervals for $\mu_{UL} - \mu_{TT}$ and $\mu_{UL} - \mu_{RL}$ both contain 0, and thus the observed differences $\bar{X}_{UL} - \bar{X}_{TT}$ and $\bar{X}_{UL} - \bar{X}_{RL}$ are not statistically significant at the family-wise significance level of 5%. The interval for $\mu_{RL} - \mu_{TT}$ did not contain 0, and thus the $\bar{X}_{RL} - \bar{X}_{TT}$ difference is statistically significant. (The interval for $\mu_{UL} - \mu_{TT}$ just barely contains 0. There is some evidence of a true difference here, but it is not statistically significant at the given family-wise significance level.)

(e) The Bonferroni intervals have the same midpoints ($\bar{X}_i - \bar{X}_j$) as the LSD intervals, but are wider (the margin of error is greater). The LSD intervals used a t value of 1.9796, whereas the Bonferroni intervals used a t value of 2.4274. (The ratio $\frac{2.4274}{1.9796} = 1.23$ shows that the Bonferroni intervals are about 23% wider than the LSD intervals.) In this specific example, this resulted in one of the comparisons (unrehearsed lying vs truth telling) changing from a significant difference under LSD to a not significant difference under Bonferroni.

(f) From the ANOVA, there is statistically significant evidence (p-value $= 0.019 < 0.05$) that the different lie conditions do not all result in the same true mean eye movement. Bonferroni confidence intervals with a 95% family-wise confidence level showed that the rehearsed lying group had a mean eye movement that was significantly greater than that of the truth tellers, but the other two differences were not statistically significant.

14.4.4 The Tukey Procedure

11. (a) The Tukey procedure is an exact procedure if the normality assumption is met, the common variance assumption is met, and the sample sizes are equal. Although not an exact procedure if the sample sizes are different, it can still be used as an approximate procedure. Here, the sample sizes are very close (41, 42, 42), and so it is very reasonable to use the Tukey procedure.

    (b) The interval for $\mu_{UL} - \mu_{TT}$ is: $63 - 42 \pm 2.3725 \times \sqrt{1561}\sqrt{1/42 + 1/41}$.
    The interval for $\mu_{RL} - \mu_{TT}$ is: $64 - 42 \pm 2.3725 \times \sqrt{1561}\sqrt{1/42 + 1/41}$.
    The interval for $\mu_{UL} - \mu_{RL}$ is: $63 - 64 \pm 2.3725 \times \sqrt{1561}\sqrt{1/42 + 1/42}$.

    (c) The interval for $\mu_{UL} - \mu_{RL}$ contains 0, and so there is not statistically significant evidence of a difference in mean eye movement between the unrehearsed lying and the rehearsed lying groups.
    The intervals for $\mu_{UL} - \mu_{TT}$ and $\mu_{RL} - \mu_{TT}$ do not contain 0, and thus there is statistically significant evidence that μ_{UL} and μ_{RL} both differ from μ_{TT}. Since both intervals lie entirely to the *right* of 0, we have strong evidence that μ_{UL} and μ_{RL} are both *greater* than μ_{TT}.

    (d) The LSD intervals used a multiplier ($t_{0.025}$) of 1.9796, the Bonferroni intervals used a multiplier ($t_{0.00833}$) of 2.4274, and the Tukey intervals used a multiplier of 2.3725. The ratio $\frac{2.3725}{1.9796} = 1.198$ shows that the Tukey intervals are about 20% wider than the LSD intervals. The ratio $\frac{2.3725}{2.4274} = 0.977$ shows that the Tukey intervals are a little narrower ($\approx 2\%$ narrower) than the Bonferroni intervals. Like the Bonferroni method, the Tukey method maintains the family-wise confidence level of 95%, but it manages to do so with narrower intervals. In this example, that resulted in one comparison crossing over to statistical significance at the 5% level. (The unrehearsed lying − rehearsed lying difference was statistically significant using the Tukey method, but not significant using the Bonferroni approach.)

    (e) From the ANOVA, there is statistically significant evidence (p-value $= 0.019 < 0.05$) that the different lie conditions do not all result in the same true mean eye movement. Tukey confidence intervals with a family-wise confidence level of 95% showed that both the unrehearsed lying group and the rehearsed lying group had mean eye movements that were significantly greater than the truth tellers, but there was not a statistically significant difference between the rehearsed and unrehearsed liars.

14.5 Examples

14.6 A Few More Points

12. We assume that we have simple random samples from the k populations, and that the populations are normally distributed with common variance σ^2. If the assumptions are not true, then our conclusions may be misleading (for example, the stated Type I error may be very different from the true probability of rejecting a true null hypothesis).

14.7 Chapter Exercises

14.8 Basic Calculations

13. (a) 0.025.
    (b) 0.95.
    (c) 6.16.
    (d) 2.93.

14.9 Concepts

14. No. In one-way ANOVA, only large values of the F test statistic give strong evidence against the null hypothesis.

15. (a) The test statistic will equal 0 if $MST = 0$ (provided $MSE \neq 0$). MST will equal 0 when there is no variability in the sample means—when the k groups all have the same value of the sample mean.
    (b) The test statistic will equal 1 if $MST = MSE$.

16. (a) If we performed a one-way ANOVA on these 3 samples, then the p-value would be small. *True. Visually there is very, very strong evidence against H_0, implying a small p-value .*
    (b) If we carried out a t-test of H_0: $\mu_A = \mu_C$ against a two-sided alternative, then the p-value would be small. *True. Visually there is very, very strong evidence against H_0, implying a small p-value .*
    (c) The use of the ANOVA procedures for the test of H_0: $\mu_A = \mu_B = \mu_C$ would be a bad idea, as the assumptions are clearly violated. *False. There is no*

visual evidence against the assumptions of normality and a common population variance.

(d) If we carried out a pooled-variance t-test of H_0: $\mu_A = \mu_C$ against a two-sided alternative, then a one-way ANOVA of the test of H_0: $\mu_A = \mu_C$, then the test statistics would be related by: $t^2 = F$. *True. For two-sample problems the pooled-variance t procedure and one-way ANOVA are equivalent tests, with* $t^2 = F$.

(e) If we carried out a pooled-variance t-test of H_0: $\mu_A = \mu_C$ against a two-sided alternative, then a one-way ANOVA of the test of H_0: $\mu_A = \mu_C$, the p-values would be exactly equal. *True. For two-sample problems the pooled-variance t procedure and one-way ANOVA are equivalent tests, resulting in the exact same p-value.*

17. (a) The test statistic will have an F distribution with 4 df in the numerator, and 45 df in the denominator. If the null hypothesis and the assumptions are true, the F test statistic has an F distribution with $k-1$ degrees of freedom in the numerator and $n-k$ degrees of freedom in the denominator. If there are 10 observations in each of 5 groups, there are $5-1=4$ degrees of freedom in the numerator, and $50-5=45$ degrees of freedom in the denominator.

(b) If the null hypothesis and the assumptions are true, then the p-value will have a uniform distribution between 0 and 1.

(c) Since the p-value is uniformly distributed between 0 and 1, the p-value will equal 0.5 on average (if the null hypothesis and the assumptions are true).

18. (a) If the null hypothesis (and the assumptions) are true, then the test statistic in one-way ANOVA has an F distribution. *True.*

(b) In one-way ANOVA, we assume that the observations within each group are normally distributed, and that all groups have the same population variance. *True.*

(c) In one-way ANOVA, we assume that the observations within each group are normally distributed, and that all groups have the same population mean. *False*

(d) The test statistic in one-way ANOVA can be negative. *False. The F statistic is a ratio of variances, and as such cannot be negative.*

(e) If the null hypothesis is false, then MST will tend to be bigger than MSE. *True.*

19. (a) If the null hypothesis is true, then the F statistic will be infinite. *False (of course).*

(b) If the null hypothesis is true, then the F statistic will equal 1. *False. If* H_0 *is true, the F statistic will have a median that is near 1, but it can take on any value between 0 and* ∞.

(c) If the null hypothesis is false, then the F statistic will sometimes be less than one. *True. The F statistic can take on any nonnegative value, regardless of whether the null hypothesis is true or false.*

(d) If the null hypothesis is false, then the F statistic will sometimes be less than 0. *False. The F statistic cannot be negative.*

(e) If the null hypothesis is false, then the p-value will be less than 0.05. *False. If the null hypothesis is false, we can still end up with a large p-value.*

20. (a) If $MSE = 0$, then the p-value will equal the F statistic. *False.*

(b) If the population means are equal, then the F statistic will equal 1. *False. If the population means are equal, then H_0 is true and the F statistic will have an F distribution.*

(c) If the sample means are all equal, then the F statistic will equal 1. *False. If the sample means are equal, the F statistic will equal 0.*

(d) If the sample means are all equal, then the null hypothesis is true. *False. If the sample means are all equal, then there is absolutely no evidence whatsoever against H_0. But that still does not mean it must be true.*

(e) The mean square error is the pooled variance. *True.*

21. (a) If the population means are very different, we expect the F statistic to be greater than one. *True. If H_0 is true (the population means are equal), then the F statistic will have a median that is somewhere around 1. If H_0 is false, the F statistic will tend to be larger than 1.*

(b) We will reject the null hypothesis at the 5% level whenever the sample means are all equal. *False. When the sample means are all equal there is absolutely no evidence whatsoever against the null hypothesis.*

(c) An assumption of one-way ANOVA is that the groups all have different population variances. *False. One-way ANOVA assumes the population variances are equal.*

(d) If the assumptions of the pooled-variance t procedures are met, then the t test and the F test are equivalent tests, with $t^2 = F$. *True.*

(e) The equal variance assumption is not important for large sample sizes, due to the central limit theorem. *False. A violation of the equal variance assumption can be very problematic, even for large sample sizes. The central limit theorem can help out with a violation of the normality assumption.*

14.10 Applications

22. (a) The ANOVA assumptions appear to be reasonable here. There are a few outliers, which can cause problems sometimes, but there are not many and they are not that extreme. The standard deviations between the groups are very similar, so the assumption of equal population variances appears to be

reasonable. The boxplots do not give any indication of non-normality. (And the normality assumption is not very important, since the sample sizes are quite large.)

(b) There is extremely strong evidence (p-value $< 2.2 \times 10^{-16}$) that the different treatments do not all result in the same mean lifetime. (In other words, there is very strong evidence that the diets do not all have the same effect.)

(c) There are 6 groups, so there are $\binom{6}{2} = 15$ pairwise comparisons ($\mu_{NP} - \mu_{R1}$, $\mu_{NP} - \mu_{R2}$, $\ldots, \mu_{R4} - \mu_{R5}$).

(d) A 95% confidence interval for $\mu_{NP} - \mu_{R1}$ is $\bar{X}_{NP} - \bar{X}_{R1} \pm t_{.025} SE(\bar{X}_{NP} - \bar{X}_{R1})$, where $SE(\bar{X}_{NP} - \bar{X}_{R1}) = \sqrt{MSE}\sqrt{1/n_{NP} + 1/n_{R1}} = \sqrt{44.5}\sqrt{1/49 + 1/57} = 1.2996$. The t has 325 degrees of freedom (the degrees of freedom for error), yielding $t_{.025} = 1.9673$ (found using software – the value may be slightly different if you look it up in a t table). The interval is thus:

$$\begin{gathered}\bar{X}_{NP} - \bar{X}_{R1} \pm t_{.025} SE(\bar{X}_{NP} - \bar{X}_{R1}) \\ 27.4 - 32.7 \pm 1.9673 \times 1.2996 \\ -5.3 \pm 2.56\end{gathered}$$

which works out to $(-7.9, -2.7)$.
The interval lies entirely to the left of 0, and thus there is a statistically significant difference between $\bar{X}_{NP}$ and $\bar{X}_{R1}$. (There is strong evidence that μ_{NP} is less than μ_{R1}.)

Non-technical summary: It appears as though mice assigned to the calorie-restricted diet R1 tend to live longer than those without calorie restriction.

(e) A 95% confidence interval for $\mu_{R2} - \mu_{R3}$ is $\bar{X}_{R2} - \bar{X}_{R3} \pm t_{.025} SE(\bar{X}_{R2} - \bar{X}_{R3})$, where $SE(\bar{X}_{R2} - \bar{X}_{R3}) = \sqrt{MSE}\sqrt{1/n_{R2} + 1/n_{R3}} = \sqrt{44.5}\sqrt{1/71 + 1/56} = 1.1922$. The t has 325 degrees of freedom (the degrees of freedom for error), yielding $t_{.025} = 1.9673$. The interval is thus:

$$\begin{gathered}\bar{X}_{R2} - \bar{X}_{R3} \pm t_{.025} SE(\bar{X}_{R2} - \bar{X}_{R3}) \\ 42.3 - 42.9 \pm 1.9673 \times 1.1922 \\ -0.6 \pm 2.3454\end{gathered}$$

which works out to $(-2.9, 1.7)$.
The interval contains 0, and thus there is not a statistically significant difference between $\bar{X}_{R2}$ and $\bar{X}_{R3}$. (There is not significant evidence that μ_{R2} differs from μ_{R3}.)

Non-technical summary: It appears as though mice assigned to the R2 diet may have the same or similar mean lifetime as those on the R3 diet.

23. (a) There are no outliers or major skewness (nothing indicating non-normality), which is good. The control group has a much greater sample standard devia-

tion than the antivenom groups (113 vs 56 and 45). This is a little troubling, as the ANOVA assumes equal population variances. However, the sample sizes are similar, and so the effect of different variability may not be that great. It is reasonable to use ANOVA, but it is not the ideal situation.

(b) There is absolutely no evidence (p-value = 0.98) of a difference in the population means of the change in leg volume. There is no evidence that the antivenom had any effect on the mean swelling.

(c) Since we had no evidence against the null hypothesis in the ANOVA, the analysis is complete. (There is no need to carry out pairwise comparisons. The researchers could, if they so desired, compare the two antivenom groups with a t test, since they had this comparison in mind before the experiment, but there would not be a significant difference in that comparison.)

24. (a) The CBT group shows several outliers, which can be a concern, but they are not that extreme and it may be the result of having a small IQR. Other than the outliers, the ANOVA assumptions are well satisfied—the distributions look roughly symmetric and the standard deviations are very similar. It's reasonable to use ANOVA here.

(b) There is very strong evidence (p-value = 0.006) that the groups do not all have the same population mean change in weight. (There is very strong evidence that the treatments do not all have the same effect.)

(c) There are $\binom{3}{2} = 3$ pairwise comparisons ($\mu_C - \mu_{CBT}$, $\mu_C - \mu_{FT}$, $\mu_{CBT} - \mu_{FT}$).

(d) A 95% confidence interval for $\mu_i - \mu_j$ is $\bar{X}_i - \bar{X}_j \pm t_{.025} SE(\bar{X}_i - \bar{X}_j)$, where $SE(\bar{X}_i - \bar{X}_j) = \sqrt{MSE}\sqrt{1/n_i + 1/n_j}$. The t has 69 degrees of freedom (the degrees of freedom for error), yielding $t_{.025} = 1.995$ (found using software – the value may be slightly different if you look it up in a t table).
The interval for $\mu_C - \mu_{CBT}$ is:

$$\begin{gathered}\bar{X}_C - \bar{X}_{CBT} \pm t_{.025} SE(\bar{X}_C - \bar{X}_{CBT}) \\ -0.20 - 1.36 \pm 1.995 \times \sqrt{11.66}\sqrt{1/26 + 1/29} \\ -1.56 \pm 1.8399\end{gathered}$$

which works out to $(-3.40, 0.28)$.
The interval for $\mu_C - \mu_{FT}$ is:

$$\begin{gathered}\bar{X}_C - \bar{X}_{FT} \pm t_{.025} SE(\bar{X}_C - \bar{X}_{FT}) \\ -0.20 - 3.30 \pm 1.995 \times \sqrt{11.66}\sqrt{1/26 + 1/17} \\ -3.50 \pm 2.1248\end{gathered}$$

which works out to $(-5.62, -1.38)$.

The interval for $\mu_{CBT} - \mu_{FT}$ is:

$$\begin{gathered}\bar{X}_{CBT} - \bar{X}_{FT} \pm t_{.025} SE(\bar{X}_{CBT} - \bar{X}_{FT}) \\ 1.36 - 3.30 \pm 1.995 \times \sqrt{11.66}\sqrt{1/29 + 1/17} \\ -1.94 \pm 2.0809\end{gathered}$$

which works out to $(-4.02, 0.14)$.

Only the $\bar{X}_C - \bar{X}_{FT}$ comparison has a statistically significant difference at the 5% level. (There is statistically significant evidence that μ_C and μ_{FT} differ.)

Non-technical summary: The women in the Family Therapy treatment group tended to gain more weight than the women in the control group. There was not strong evidence of a difference in weight gain between the control and CBT groups, or the CBT and FT groups.

(e) There are 3 pairwise comparisons. To ensure the family-wise confidence level is at least 95%, each interval should be done at a confidence level of $(1 - \frac{0.05}{3})100\% = 98.33\%$ ($\alpha' = 0.05/3 = 0.0167$). We need to find $t_{0.0167/2} = t_{0.0083}$ with 69 degrees of freedom. Using software, we can find that with 69 degrees of freedom $t_{0.0083} = 2.4537$.

The interval for $\mu_C - \mu_{CBT}$ is:

$$\begin{gathered}\bar{X}_C - \bar{X}_{CBT} \pm t_{.025} SE(\bar{X}_C - \bar{X}_{CBT}) \\ -0.20 - 1.36 \pm 2.4537 \times \sqrt{11.66}\sqrt{1/26 + 1/29} \\ -1.56 \pm 2.2629\end{gathered}$$

which works out to $(-3.82, 0.70)$.

The interval for $\mu_C - \mu_{FT}$ is:

$$\begin{gathered}\bar{X}_C - \bar{X}_{FT} \pm t_{.025} SE(\bar{X}_C - \bar{X}_{FT}) \\ -0.20 - 3.30 \pm 2.4537 \times \sqrt{11.66}\sqrt{1/26 + 1/17} \\ -3.50 \pm 2.6133\end{gathered}$$

which works out to $(-6.11, -0.89)$.

The interval for $\mu_{CBT} - \mu_{FT}$ is:

$$\begin{gathered}\bar{X}_{CBT} - \bar{X}_{FT} \pm t_{.025} SE(\bar{X}_{CBT} - \bar{X}_{FT}) \\ 1.36 - 3.30 \pm 2.4537 \times \sqrt{11.66}\sqrt{1/29 + 1/17} \\ -1.94 \pm 2.5593\end{gathered}$$

which works out to $(-4.50, 0.62)$.

Only the $\bar{X}_C - \bar{X}_{FT}$ comparison has a statistically significant difference at a family-wise significance level of 5%. (There is statistically significant evidence that μ_C and μ_{FT} differ.) In this case the conclusions from the LSD and the

Bonferroni procedures are essentially the same (but the Bonferroni procedure results in slightly wider intervals).
Non-technical summary: The women in the Family Therapy treatment group tended to gain more weight than the women in the control group. There was not strong evidence of a difference in weight gain between the control and CBT groups, or the CBT and FT groups.

25. (a) The standard deviations are very similar, so there is no issue with the common variance assumption. Overall, ANOVA is likely reasonable here. But we should plot the data, as this can give us an overall view of what we are dealing with, help to assess if the normality assumption is reasonable, and show any outliers or other problems.
    (b) There is no evidence (p-value $= 0.64$) of a difference in the true mean SDO of the different ethnic groups. (These conclusions apply only to the male inmate population from which the sample was drawn.)

26. (a) Any subsequent analysis would depend on the assumption that it is reasonable to pool the individuals together. If the distribution of SDO did vary between the ethnic groups in a meaningful way, then conclusions based on a pooled analysis would be suspect. For example, if the pooled sample showed no relationship between age and SDO, this might be because the effect of age differed between the ethnic groups and the pooling averaged it out. If we pool over an important variable, the resulting conclusions can sometimes be very misleading.
    (b) There is extremely strong evidence (p-value $= 3.8 \times 10^{-8}$) that the age groups do not all have the same population mean SDO. (These conclusions apply only to the male inmate population from which the sample was drawn.)
    (c) There is extremely strong evidence that there is a relationship between age and SDO. From the values given in the table, it appears as though younger inmates tend to have a higher SDO, and are thus more accepting of the notion that it is reasonable for some groups to dominate others. (We could investigate this a little more formally by carrying out the $\binom{4}{2} = 6$ pairwise comparisons. And since age is a quantitative variable, we could also investigate this by using *regression* techniques instead of ANOVA.)

27. (a) The ANOVA table is given in Table 14.1.

Source	DF	SS	MS	F	p-value
Groups	$6 - 1 = 5$	$8117.2 - 6862.8 = 1254.4$	$\frac{1254.4}{5} = 250.88$	$\frac{250.88}{60.2} = 4.167$	$< .01$
Error	$120 - 6 = 114$	$60.2 \times 114 = 6862.8$	60.2	—	—
Total	$120 - 1 = 119$	8117.2	—	—	—

Table 14.1: ANOVA table for the teaching method question.

(b) $F = 4.167$. The p-value is the area to the right of 4.167 under an F distribution with 5 and 114 degrees of freedom. Using a computer: p-value $= 0.0016$. Using the table: $0.001 < p$-value < 0.01.

(c) There is significant evidence that the teaching methods do not all result in the same mean test performance.

(d) $SE(\bar{X}_i - \bar{X}_j) = \sqrt{60.2}\sqrt{1/20 + 1/20} = 2.4536$. We now need $t_{.05}$ with 114 degrees of freedom. The correct value is approximately 1.658 (by computer, it might be slightly different using the table), yielding a margin of error of $1.658 \times 2.4536 = 4.068$.

(e) We require the different groups to have normally distributed populations with common variance σ^2. Normal quantile-quantile plots can help us check the normality assumption. Summary statistics and boxplots could help us determine if the equal variance assumption is reasonable.

28. (a)

Source	DF	SS	MS	F
Treatments	$9 - 1 = 8$	$612 - 324 = 288$	$\frac{288}{8} = 36$	$\frac{36}{12} = 3.00$
Error	$36 - 9 = 27$	$27 * 12 = 324$	12	
Total	$36 - 1 = 35$	612		

(b) Here we need the area to the right of 3 under an F distribution with 8 and 27 degrees of freedom. Using a computer: p-value $= 0.015$. Using the table: $0.01 < p$-value < 0.025.

(c) There is significant evidence against the null hypothesis, which means there is significant evidence that not all of the seat belt types have the same population mean breaking strength.

(d) $2.052 \times \sqrt{12}\sqrt{1/4 + 1/4} = 5.026$

(e) i. The sample means were not all equal. *True. If the sample means were all equal, then $MST = 0$.*

ii. This was a balanced experimental design (same sample size in each group). *True, there were 4 observations in each group.*

iii. We know for certain the population mean breaking strengths of the different groups are not all equal. *False. Statistical inference is not about being certain.*

iv. The pooled sample variance is equal to 12. *True (the pooled variance is MSE).*

29. (a) H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (the mean fuel use is the same for all 5 additives). H_a: $\mu_i \neq \mu_j$ for at least one i, j combination. (the means of the 5 additives are not all equal).

(b) We require the different groups to have normally distributed populations with common variance σ^2. Normal quantile-quantile plots can help us check the normality assumption (although these may not be very informative with the small sample sizes). Summary statistics and boxplots could help us determine if the equal variance assumption is reasonable.

(c) The ANOVA table is given in Table 14.2.

Source	DF	SS	MS	F	p-value
Groups	$5-1=4$	680.2	$\frac{680.2}{4}=170.05$	$\frac{170.05}{18.533}=9.175$	p-value $< .001$
Error	$20-5=15$	$958.2-680.2=278$	18.533	—	—
Total	$20-1=19$	958.2	—	—	—

Table 14.2: ANOVA table for the fuel additives question.

(d) $F = 9.175$. The p-value is the area to the right of 9.175 under an F distribution with 4 and 15 degrees of freedom. Using a computer: p-value $= 0.0006$. Using the table: p-value < 0.001.

(e) There is very strong evidence (p-value $= 0.0006$) that the additives do not all result in the same population mean fuel use.

30. (a) H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (the population mean strength is the same for all 4 solder types).
H_a: $\mu_i \neq \mu_j$ for at least one i, j combination. (the population means of the 4 solder types are not all equal).

(b) There is very strong evidence (p-value $= 0.00017$) that the types of tin-lead solder do not all result in the same mean shear strength.

(c) $2.120 \times \sqrt{4.425}\sqrt{1/5 + 1/5} = 2.820$.

Chapter 15

Introduction to Simple Linear Regression

J.B.'s strongly suggested exercises: 2, 4, 6, 8, 9, 12, 14, 15, 16, 18, 20, 37, 40, 41, 42, 43

15.1 Introduction

1. In regression analysis, we use one or more explanatory variables to help us predict or estimate a quantitative response variable. For example, in regression analysis we might predict a person's resting heart rate based on a number of variables, such as age, weight, sex, and various health indicators.

2. (a) Since we wish to use high school GPA to help predict university GPA, high school GPA will be the *explanatory* variable and university GPA will be the *response* variable.
   (b) There are a number of variables that might help to predict university GPA, including the student's socioeconomic background, gender, and high school they attended. Regression analysis can help us to determine which of these are useful predictors.

3. In simple linear regression there is one explanatory variable (X). In multiple linear regression there is more than one explanatory variable ($X_1, X_2, \ldots$). (In both cases there is a single response variable Y.)

15.2 The Linear Regression Model

4. The terms in the model:
   - Y is the response variable. (We may wish to predict the value of Y.)
   - X is the explanatory variable. (We use the explanatory variable X to help predict the value of the response variable Y.)
   - β_0 is the Y intercept. In other words, it is the theoretical mean of Y when $X = 0$. Although the intercept is required for the line, it may not have much practical meaning.
   - β_1 is the slope of the line. In other words, it is the change in the mean of Y for a 1 unit increase in X. (But note that there is not necessarily a causal link between X and Y.)
   - ϵ is a random error term, representing the fact that the response variable Y varies about the regression line.

5. $\hat{\beta}_1$ is a statistic that is calculated using sample data. It estimates β_1, the true slope (a parameter).
   $\hat{\beta}_0$ is a statistic that is calculated using sample data. It estimates β_0, the true intercept (a parameter).

6. $\hat{\beta}_0 \approx 10$ and $\hat{\beta}_1 \approx -2$.

   $\hat{\beta}_1$ is the slope. We know from our grade school days that Slope $= \frac{\text{Rise}}{\text{Run}}$. When $X = 2$, Y is approximately 6, and when $X = 10$, Y is approximately -10. This is a drop of approximately 16 units (a *rise* of approximately -16) over a run of 8. $\hat{\beta}_1 = \frac{\text{Rise}}{\text{Run}} \approx \frac{-16}{8} = -2$. (The slope of the given line is approximately -2.) $\hat{\beta}_0$ is the value of $\hat{Y}$ when $X = 0$. At $X = 2$, $\hat{Y} \approx 6$. With a slope of approximately -2, this implies that at $X = 0$, $\hat{Y} \approx 10$, so $\hat{\beta}_0 \approx 10$.

15.3 The Least Squares Regression Model

7. We use the least squares method—the parameter estimate are chosen such that $\sum e_i^2 = \sum(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$ is a minimum. In other words, the parameter estimates are chosen such that the sum of the squared vertical distances from the points to the line is a minimum.

8. (a) $\hat{\beta}_1 = \frac{SP_{XY}}{SS_{XX}} = \frac{10.1}{20.2} = 0.5$. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 23.0 - 0.5 \times 12.0 = 17.0$.
   (b) The units of $\hat{\beta}_0$ are metres (the same units as Y). The units of $\hat{\beta}_1$ are metres/second ($\frac{\text{Units of } Y}{\text{Units of } X}$).

9. (a) $\hat{Y} = -5.73529 + 0.63554X$.
   (b) $\hat{\beta}_1 = 0.63554$ is the estimated change in the mean abdomen length for a 1 mm increase in snout-vent length. It is also the predicted difference in abdomen length between two lizards that differ by 1 mm in their snout-vent length. For example, we would predict that a lizard with a snout-vent length of 51 mm would have an abdomen length that is approximately 0.64 mm greater than that of a lizard with a snout-vent length of 50 mm.
   (c) $\hat{\beta}_0 = -5.73529$ is the estimated mean abdomen length for a lizard with a snout-vent length of 0 mm. Note that this makes absolutely no sense from a practical point of view. The snout-vent length of 0 is far beyond the range of the observed data, and not something of any practical interest, so in this example the value of the intercept has no practical meaning beyond being a necessary part of the line.
   (d) $\hat{Y} = -5.73529 + 0.63554 \times 52.0 = 27.31279$.
   (e) In Question 9d we found that the predicted abdomen length for a lizard with a snout-vent length of 52.0 mm is $\hat{Y} = -5.73529 + 0.63554 \times 52.0 = 27.31279$ mm. This lizard has a residual of: $Y - \hat{Y} = 27.3 - 27.31279 = -0.01279$.
   (f) 0, since the residuals always sum to 0 in linear regression.

15.4 Statistical Inference in Simple Linear Regression

15.4.1 Model Assumptions

10. We assume that the observations are independent, and that the ϵ term is a random variable that is normally distributed with a mean of 0 and the same variance (σ^2) at every value of X. Under these assumptions, $Y \sim N(\beta_0 + \beta_1 X, \sigma^2)$.

11. (a) $s^2 = \dfrac{\sum e_i^2}{n-2} = \dfrac{3.1^2 + 2.4^2 + (-6.1)^2 + (-.5)^2 + 8.9^2 + (-12.6)^2 + (4.8)^2}{7-2} =$ 62.768.
    (b) The sum of the residuals is 0. The residuals always sum to 0 in a least squares regression.

15.4.2 Statistical Inference for the Parameter β_1

12. (a) Never. We carry out hypothesis tests on *parameters*, not *statistics*.
    (b) Rarely. This might be a point of interest for an investigator from time to time, but not often.

(c) We test H_0: $\beta_1 = 0$ *very* frequently in linear regression. It is often one of the main points of interest in a study (since rejecting this null hypothesis indicates strong evidence of a relationship between the explanatory and response variables).

(d) Almost never. This hypothesis has no practical meaning in essentially all situations.

13. $\hat{\beta}_1 \sim N(\beta_1, \dfrac{\sigma^2}{SS_{XX}})$

14. (a) There are $n-2 = 22-2 = 20$ degrees of freedom. With 20 degrees of freedom, $t_{.025} = 2.086$. A 95% confidence interval for β_1 is given by

$$\hat{\beta}_1 \pm t_{.025} SE(\hat{\beta}_1)$$
$$0.63554 \pm 2.086 \times 0.04847$$
$$0.63554 \pm 0.1011$$

which works out to (0.534, 0.737).

(b) We can be 95% confident that β_1 lies between 0.534 and 0.737, where β_1 is the change in the true mean abdomen length corresponding to a 1 mm increase in snout-vent length.

(c) H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$. $t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{0.63554-0}{0.04847} = 13.11$. (Note that the t test statistic is given in the output.) The p-value is double the area to the right of 13.11 under a t distribution with 20 degrees of freedom. The p-value is 2.79×10^{-11} (given in the output). (If this was not available in the output, and we had to use a t table, we could say only that the p-value is very, very close to 0.) There is extremely strong evidence that the true slope β_1 is greater than 0. This implies that as snout-vent length increases, abdomen length tends to increase. (Which is hardly surprising.)

(d) H_0: $\beta_0 = 0$, H_a: $\beta_0 \neq 0$. $t = \frac{\hat{\beta}_0 - 0}{SE(\hat{\beta}_0)} = \frac{-5.73529-0}{2.50608} = -2.289$. (Note that the t test statistic is given in the output.) The p-value is double the area to the left of -2.289 under a t distribution with 20 degrees of freedom. The p-value is 0.0331 (given in the output). (If this was not available in the output, and we had to use a t table, we could give only a range, such as $0.02 < p\text{-value} < 0.05$.) The p-value is less than the given significance level (p-value < 0.05), and so there is statistically significant evidence that the true intercept differs from 0. (It appears as though the true intercept is in fact less than 0.)

(e) There are $n-2 = 22-2 = 20$ degrees of freedom. With 20 degrees of freedom, $t_{.025} = 2.086$. A 95% confidence interval for β_0 is given by

$$\hat{\beta}_0 \pm t_{.025} SE(\hat{\beta}_0)$$
$$-5.73529 \pm 2.086 \times 2.50608$$
$$-5.73529 \pm 5.2277$$

which works out to $(-10.963, -0.507)$. Note that this 95% confidence interval lies entirely to the left of 0, which should not be surprising given the results of the hypothesis test.

15.5 Checking Model Assumptions with Residual Plots

15. Plot A shows curvature in the residuals (the residuals tend to be too small for large and small values of X, and too large near the mean of X). Our simple linear regression model is not adequate for this data.

    Plot B shows that the variance of the residuals increases as X increases. This is a problem for our model. (In these situations, the slope and intercept estimators are still unbiased estimators of their respective parameters, but our standard errors will be off.)

    Plot C shows a random scattering of points, and no problems with the assumed linear regression model.

    Plot D shows a normal quantile-quantile plot in which the points lie very close to a straight line. It indicates that the observed residuals are approximately normally distributed. This indicates that the normality assumption is reasonable.

16. The plots look good, and do not indicate any problems with the assumed model. The plot of the residuals vs snout-vent length shows a random scattering of points, without any problems like curvature or increasing variance. The points on the normal quantile-quantile plot fall roughly on a straight line. (There are 3 residuals that are a little smaller than would be expected (very mild outliers), but they are unlikely to have a major impact.) The assumptions of the model are reasonable, and the inference procedures we carried out in Question 14 are valid.

15.6 The Correlation Coefficient and the Coefficient of Determination

17. (a) $r = \frac{SP_{XY}}{\sqrt{SS_{XX}SS_{YY}}} = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}} = \frac{10.1}{\sqrt{20.2 \cdot 6.5}} = 0.8814.$
    (b) No units. (The correlation coefficient r is a unitless quantity.)

18. (a) Yes, the correlation coefficient would be a very reasonable measure of the strength of the relationship in each of the four plots, since all of the relationships look roughly linear.

(b) From software, the calculated values of the correlation coefficients: Plot A: $r = 0.523$, Plot B: $r = -0.988$, Plot C: $r = -0.884$, Plot D: $r = 0.082$. Ranking in terms of increasing values of the correlation coefficient r: Plot B < Plot C < Plot D < Plot A.

(c) Ranking in terms of increasing strength of the linear relationship:
Plot D < Plot A < Plot C < Plot B. (The greater the *magnitude* of r, the stronger the linear relationship.)

(d) A change of the scaling of X and/or Y would not change the value of the correlation coefficient r. Multiplying X or Y by a constant, or adding a constant, does not change r.

19. (a) Yes, the coefficient of determination would be a very reasonable measure of the strength of the relationship in each of the four plots, since all of the relationships look roughly linear.

(b) From software, the calculated values of the coefficient of determination: Plot A: $R^2 = 0.274$, Plot B: $R^2 = 0.977$, Plot C: $R^2 = 0.781$, Plot D: $R^2 = 0.007$. Ranking in terms of increasing values of the coefficient of determination:
Plot D < Plot A < Plot C < Plot B.

20. (a) The output shows that $R^2 = 0.8958$.

(b) Approximately 90% of the variance in abdomen length can be explained by the linear relationship with snout-vent length.

(c) The *magnitude* of r is $\sqrt{R^2} = \sqrt{0.8958} = 0.946$. r always has the same sign as the slope $\hat{\beta}_1$, and since $\hat{\beta}_1$ is positive for this data, $r = 0.946$.

15.7 Estimation and Prediction Using the Fitted Line

21. (a) $\hat{\mu}_{Y|X^*} = -5.73529 + 0.63554 \times 50.0 = 26.042$ mm.

(b) $\hat{Y} = -5.73529 + 0.63554 \times 50.0 = 26.042$ mm. (The predicted value is the same as the estimated mean.)

(c) With 20 degrees of freedom, $t_{.025} = 2.086$. The 95% interval is:

$$\hat{\mu}_{Y|X^*} \pm t_{\alpha/2} SE(\hat{\mu}_{Y|X^*})$$
$$26.042 \pm 2.086 \times 0.15447$$
$$26.042 \pm 0.322$$

Which works out to (25.72, 26.36).

(d) The standard error of the estimated mean when $X = 45$ would be *greater* than the standard error when $X = 50$, since 45 is further from the mean of X than 50 is ($\bar{X} = 51.6$.) Since the standard error is greater when $X = 45$, the confidence interval would be wider.

(e) A prediction interval for a single value is wider than the corresponding confidence interval for the mean. (There is greater uncertainty when we are predicting a single value than when we are estimating the mean.)

15.8 Transformations

15.9 A Complete Example

15.10 Outliers and Influential Points

22. (a) Points A and D have the greatest leverage, since they are furthest from the mean of X. Points B and C have very little leverage, since they are near the mean of X.
    (b) Points B and D have the residuals with the greatest magnitude. (The vertical distance from points B and D to the line is far greater than for the other two points.)
    (c) Point D would be the most influential, since it has both high leverage and falls far from the overall pattern of the observations. It would definitely have the most influence of any point in this data set.
    (d) If point D were removed, the slope would increase.

15.11 Some Cautions about Regression and Correlation

15.11.1 Always Plot Your Data

15.11.2 Avoid Extrapolating

23. (a) No, as this would be extrapolating far beyond the range of the observed data.
    (b) No, the model is not flawed. We view $\hat{\beta}_0$ as the estimated mean of Y when $X = 0$ only if $X = 0$ is within (or very near) the range of the observed data. When $X = 0$ is far beyond the range of the observed data, we typically view $\hat{\beta}_0$ simply as the Y-intercept, a necessary part of the line without any practical meaning beyond that.

15.11.3 Correlation Does Not Imply Causation

15.12 A Brief Multiple Regression Example

15.13 Chapter Exercises

15.13.1 Basic Calculations

24. (a) $e_1 = Y_1 - \hat{Y}_1$, where $Y_1 = 25$, $\hat{Y}_1 = 1.20 + 3.14(8) = 26.32$. $e_1 = 25 - 26.32 = -1.32$.
    (b) $3.14 \pm 2.306 \times 0.3436$, which works out to 3.14 ± 0.7923, or $(2.35, 3.93)$.
    (c) Yes, since the interval does not contain 0. Since the entire interval lies to the right of 0, there is strong reason to believe $\beta_1 > 0$, and there is an increasing relationship of Y with X.
    (d) $\hat{Y} = 1.20 + 3.14(120) = 378$. This prediction is an *extrapolation*, using the regression line to estimate Y for an X value that is beyond the range of the observed data. It is not a reasonable use of the regression line.
    (e) $\hat{Y} = 1.20 + 3.14(20) = 64$. This is still an extrapolation, and it's still not reasonable.

25. (a) $\hat{Y} = 4.0571 - 1.2429(1) = 2.8142$.
    (b) $e_1 = Y_1 - \hat{Y}_1 = 3.1 - 2.8142 = 0.2858$.
    (c) $-1.2429 \pm 2.353 \times .2078$, which works out to -1.2429 ± 0.4890, or $(-1.73, -0.75)$.
    (d) If there is truly no linear relationship between X and Y, then $\beta_1 = 0$. In this event, β_0 would equal the true mean of Y $(E(Y))$.
    (e) $\frac{-1.2429-2}{.2078} = -15.60587$. The two-sided p-value is double the area to the left of -15.60587 under a t distribution with $5 - 2 = 3$ degrees of freedom. This p-value is very small, p-value $< .001$. There is very strong evidence that $\beta_1 < 2$. In other words, there is very strong evidence that the true relationship between X and Y has a slope that is less than 2.
    (f) The assumptions are that the error terms (ϵ) are independent, and normally distributed with a mean of 0 and common variance. The plot doesn't show any evidence against these assumptions, but with a sample size of $n = 5$ there isn't a lot to go on.

26. (a) Yes. The two-sided p-value of the test of H_0: $\beta_1 = 0$ is .001.
    (b) $1.218 \pm 2.306 \times 0.248$, which works out to $1.218 \pm .5719$, or $(.646, 1.790)$.

15.13.2 Concepts

27. The term *extrapolating* refers to using the regression line to predict or estimate values of Y for values of X that are beyond the range of the observed data. This should be avoided.

28. (a) Plot 1 shows a strong relationship between X and Y, but it is clearly not a straight line relationship. Plots 2 and 3 look reasonable.
    (b) Plot 1 has the strongest relationship.
    (c) Plot 2 shows the strongest linear relationship.
    (d) In decreasing order of the strength of the linear relationship: Plot 2 > Plot 3 > Plot 1. Since r^2 measures the strength of the linear relationship, .9509 corresponds to Plot 2, .5935 corresponds to Plot 3, and .04392 corresponds to Plot 1.

29. No. If r is very close to 0, then the sample shows no *linear* relationship between X and Y. There could still be a strong relationship between X and Y, but not a simple straight line relationship. See, for example, the relationship in Figure 15.1.

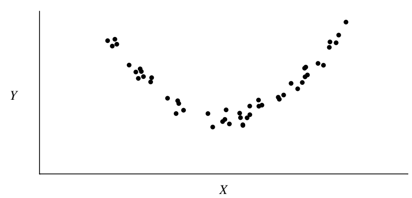


Figure 15.1: A curvilinear relationship (r is close to 0).

And if r is close to 0, there is always the chance that there is truly a strong straight line straight line relationship between X and Y, and just we happened to see a value of r close to 0.

30. An r^2 value of .95 implies a very strong linear relationship between X and Y. For example, Figure 15.2 shows a relationship with $r^2 = 0.81$. But even though the relationship is strong, there is not necessarily a causal link.

31. (a) $\hat{Y} = -1.0 + 5.1X$.
    (b) $\hat{Y} = -1.0 + 5.1(2.5) = 11.75$.
    (c) $5.10 \pm 4.303 \times .7746$, which works out to 5.10 ± 3.333, or $(1.77, 8.43)$.
    (d) $t = \frac{5.10-0}{.7746} = 6.584$.
    (e) $.02 < p\text{-value} < .05$.
    (f) The four predicted values are $\hat{Y} = -1.0+5.1(1) = 4.1$, $\hat{Y} = -1.0+5.1(2) = 9.2$,

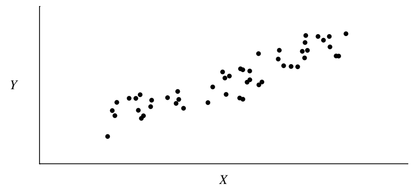


Figure 15.2: A relationship with $r^2 = 0.81$.

$\hat{Y} = -1.0 + 5.1(3) = 14.3$, $\hat{Y} = -1.0 + 5.1(4) = 19.4$. The four residuals are $4.1 - 4.1 = 0$, $8.2 - 9.2 = -1.0$, $16.3 - 14.3 = 2.0$, $18.4 - 19.4 = -1.0$. This leaves a sample variance about the regression line of

$$s^2 = \frac{0^2 + (-1)^2 + 2.0^2 + (-1.0)^2}{4-2} = \frac{6}{2} = 3$$

32. $s^2 = \frac{\sum e_i^2}{n-2}$ is the best estimate of σ^2. We aren't given the fourth residual, but we do know that the residuals must sum to 0 ($\sum e_i^2 = 0$). Since the first 3 residuals sum to 4.7, this implies the fourth residual must have been -4.7. $s^2 = \frac{2.0^2+3.4^2(-0.7)^2+(-4.7)^2}{4-2} = 19.07$.

33. (a) In reality there is absolutely no relationship between X and Y in this situation (they are randomly generated values, randomly paired together), so $\beta_1 = 0$. The null hypothesis and the assumptions of the t test are true, and so the test statistic will have a t distribution with $n - 2 = 10 - 2 = 8$ degrees of freedom.
    (b) 0, since that is the mean of a t distribution with 8 degrees of freedom.
    (c) In the event the null hypothesis is true (and the assumptions are true), the p-value has a uniform distribution between 0 and 1. so here the p-value would have a uniform distribution between 0 and 1.
    (d) 0.5 (the mean of a uniform distribution between 0 and 1).

34. The test of H_0: $\beta_1 = 0$ (the null hypothesis of no linear relationship between X and Y).

35. Since 0 lies outside of the 95% confidence interval for β_1, the two-sided p-value of the test of H_0: $\beta_1 = 0$ is less than .05.

36. (a) The method used to find the sample slope and intercept is called the method of least cubes. *False. We use the method of least squares.*
    (b) The residuals in a simple linear regression always sum to 0. *True.*
    (c) If $\bar{X} = 0$, then $\hat{\beta}_1 = 0$. *False.*
    (d) If $\bar{Y} = 0$, then $\hat{\beta}_0 = 0$. *False.*

(e) r has the same sign as $\hat{\beta}_1$. *True. The sign of r is always the same as the sign of $\hat{\beta}_1$.*

37. (a) $\hat{\beta}_0$ has the same sign as $\hat{\beta}_1$. *False. They can be equal, but they certainly don't have to be.*
    (b) If all data points fall perfectly on a line, then $r = 1$ or -1. *True.*
    (c) The least squares regression line always passes through the point (0, 0). *False. The line will only pass through the origin if $\hat{\beta}_0 = 0$.*
    (d) The least squares regression line always passes through the point $(\bar{X}, \bar{Y})$. *True. (To show this, try putting $\bar{X}$ and the formula for $\hat{\beta}_0$ into the least squares line).*
    (e) If there is no linear relationship between Y and X, then $r = 1$. *False. We can't have a stronger sample relationship than $r = 1$ (all points fall perfectly on a line when $r = 1$).*

38. (a) Tests and confidence intervals on $\hat{\beta}_0$ are the most common inference procedures in regression. *False, for a couple of reasons. First, we do not carry out inference procedures for the statistic $\hat{\beta}_0$. We may wish to carry out inference procedures for the parameter β_0, but these procedures are not usually of practical interest. Inference for β_1 is much more common.*
    (b) If the sample shows no linear relationship between Y and X (a random scattering of points, say), then r and r^2 will be close to 0. *True.*
    (c) $\epsilon_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$. *False. $e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$, not ϵ_i. $\epsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$*
    (d) The residuals (the e_i values) are uncorrelated. *False. We assume that the true error terms (the values of ϵ) are uncorrelated, but the observed residuals are correlated.*
    (e) The sample correlation coefficient is the proportion of the variance in Y that is attributable to the linear relationship with X. *False. This is the meaning of the* coefficient of determination.

39. (a) Instead of least squares, it would be better to choose a regression line based on minimizing the sum of the absolute value of the residuals, but statisticians have not yet figured out how to do so. *False. We can do this, but the least squares method has some nice mathematical properties.*
    (b) The estimated standard deviation in a simple linear regression is the sum of the squared residuals, divided by the sample size. *False.* $s^2 = \frac{\textit{Sum of squared residuals}}{n-2}$.
    (c) If the test of the null hypothesis $\beta_1 = 0$ results in a very small p-value, then r^2 must have been large. *False. Statistical significance is largely tied to sample size. A small R^2 value can still result in statistical significance, if the sample size is large enough.*
    (d) If $s = 0$ and $\hat{\beta}_1 > 0$, then $r = 1$. *True.* $s = 0$ *implies that all points fall perfectly on a line (there is no variability about the regression line), and r has the same sign as the sample slope $\hat{\beta}_1$.*

15.13.3 Applications

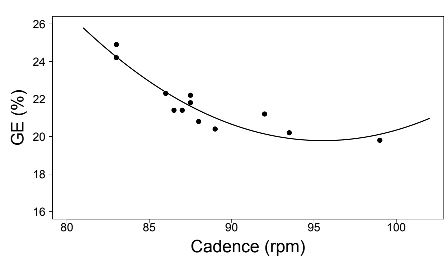
40. (a) Overall the simple linear regression model looks to be reasonable here. A straight line fit looks very reasonable (there is no indication of curvature). There does appear to be some increasing variance. (As head width increases, the variance of the residuals increases. We often see this with count data, as the variance often increases with the mean.) We do have more advanced statistical methods that deal with this type of data, but overall the simple linear regression model is not bad.
    (b) No, the assumptions of the simple linear regression model are not perfectly true in this case. The model assumes that for a given X, Y is normally distributed about the true line. That cannot possibly be true here, as Y represents a discrete count and thus cannot be perfectly normally distributed. But we can still use the model. It may not be perfect, but it may very well provide a very reasonable approximation.
    (c) $\hat{Y} = -29.400 + 17.004X$, where $\hat{Y}$ represents the predicted number of larvae and X represents head width.
    (d) A 1 mm increase in head width (X) results in an increase in the estimated mean number of larvae (Y) of 17.004.
    (e) It is estimated that the larger blowflies carry $0.2 \times 17.004 = 3.4$ more larvae on average.
    (f) Yes, there is extremely strong evidence of a relationship between female head width and the number of live larvae that they carry. The p-value of the test of H_0: $\beta_1 = 0$ is 9.01×10^{-14}. This is a minuscule p-value, and there is almost surely a relationship between these variables.
    (g) $\hat{Y} = -29.400 + 17.004 \times 4.0 = 38.616$. (If we were using the model to predict the count for a single female with a head width of 4.0 mm, our best guess would be 39 larvae.)
    (h) 0.6969 (the value of R^2 given in the output).

41. (a) H_0: $\beta_1 = 0$ (there is no linear relationship between hand grip strength and facial attractiveness).
    H_a: $\beta_1 \neq 0$ (there is a linear relationship between hand grip strength and facial attractiveness).
    $t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{0.01309}{0.01845} = 0.709$ (note that this t statistic is given in the output).
    The p-value is double the area to the right of 0.709 under a t distribution with $n - 2 = 32 - 2 = 30$ degrees of freedom. The p-value of 0.4836 is given in the output. This is a large p-value, giving no evidence against the null hypothesis. This sample gives no evidence of a relationship between hand grip strength and facial attractiveness.
    (Side note: The authors of the original paper found that after using more advanced statistical methods to control for other important variables (e.g.

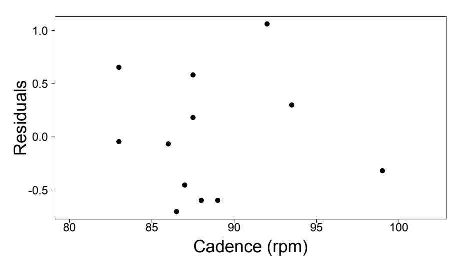
weight), there was evidence of a relationship.)

(b) 1.65%. (The value of R^2 is given in the output as 0.0165.)

42. (a) Although the sample size is small and there is nothing definitive, it appears as though the relationship between GE and cadence may be more of a curvilinear relationship. The residuals corresponding to both the small and large values of cadence are large and positive, whereas the residuals corresponding to values of cadence near the mean are negative. (The model may be *underestimating* GE for large and small cadence values, and *overestimating* GE for values of cadence near the mean.) The simple linear regression model may not be adequate.
One option for us would be to fit a curve through the points using multiple regression techniques. (We crazy statisticians can do that!) For this example, if we fit a curve by including an X^2 term in our model, we would get the curvilinear relationship shown in Figure 15.3a. The residuals for this model are shown in Figure15.3b. This model seems to fit the data much better.



(a) Scatterplot with a fitted curve.



(b) Residuals for the fitted curve.

Figure 15.3: Scatterplot of GE vs cadence and plot of the residuals for the fitted curve.

(b) Yes, there is very strong evidence of a relationship between cadence and GE. The p-value of the test of H_0: $\beta_1 = 0$ is given in the R output as 0.00130. There is strong evidence that as cadence increases, GE tends to decrease.

43. (a) H_0: $\beta_0 = 0$, H_a: $\beta_0 \neq 0$. This test does not have much practical meaning. This tests the hypothesis that the mean memory score is 0 when the right hippocampus has a volume of 0. This doesn't have any practical meaning.

(b) H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$. Yes, this is meaningful. It is a test of the null hypothesis that there is no linear relationship between right hippocampus volume and score on the memory test.

(c) There is very strong evidence (two-sided p-value $= 0.0017$) of an increasing relationship between right hippocampus volume and score on the Wechsler memory scale.

(d) $0.08835 \pm 2.086 \times 0.02440$, which works out to 0.08835 ± 0.0509 or (0.037, 0.139).

(e) $e_1 = Y_1 - \hat{Y}_1$. $Y_1 = 28$. $\hat{Y}_1 = -41.72277 + 0.08835 \times 955 = 42.65148$.

$e_1 = 28 - 42.65148 = -14.65$.

(f) 0.396 (The value of R^2 given in the output.)

(g) The magnitude of r is $\sqrt{r^2} = \sqrt{0.396} = 0.629$. And since r has the same sign as the slope, $r = 0.629$.

44. (a) $\hat{Y} = 0.13215 + 0.362 \times 44 = 16.06015$.

(b) That is a test on the slope: H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$. A p-value of 7.59×10^{-11} gives extremely strong evidence against the null hypothesis (we would see a relationship this strong less than 1 time in 100 billion if there is truly no relationship). There is extremely strong evidence that as the insulation level rises, the temperature rise above ambient tends to increase.

(c) H_0: $\beta_1 = 0.30$, H_a: $\beta_1 \neq 0.30$.

$t = \frac{0.3620-0.30}{0.03211} = 1.931$. The p-value is double the area to the right of 1.931 under a t distribution with $25 - 2 = 23$ degrees of freedom. By computer: p-value $= 0.066$. By the table: $0.05 < p\text{-value} < 0.10$. There is not significant evidence at $\alpha = 0.05$ that the true slope differs from 0.30.

45. (a) $\hat{\beta}_1 = \dfrac{SS_{XY}}{SS_{XX}} = \dfrac{1100}{917} \approx 1.1996$.

(b) $SE(\hat{\beta}_1) = \sqrt{\frac{s^2}{SS_{XX}}} = \sqrt{\frac{24.935}{917}} \approx 0.1649$.

(c) $1.1996 \pm 2.306 \times 0.1649$, which works out to 1.1996 ± 0.3803, or $(0.8193, 1.5799)$.

(d) Yes. The entire interval lies well to the right of 0.

(e) H_0: $\beta_1 = 1$, H_0: $\beta_1 \neq 1$. $t = \frac{1.1996-1}{.1649} = 1.2104$. For a two-sided alternative, $0.20 < p\text{-value} < 0.30$. We have no evidence that β_1 differs from 1.

(f) $\hat{Y} = 161.6358 + 1.1996(92) = 271.999$.

(g) $s^2 = 24.935$.

(h) Under our usual assumptions, $\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{SS_{XX}})$

(i) $\hat{Y} = 161.6358 + 1.1996(-2) = 159.2366$. This is a nonsensical estimate, as it estimates the height of a tree that is -2 months old.