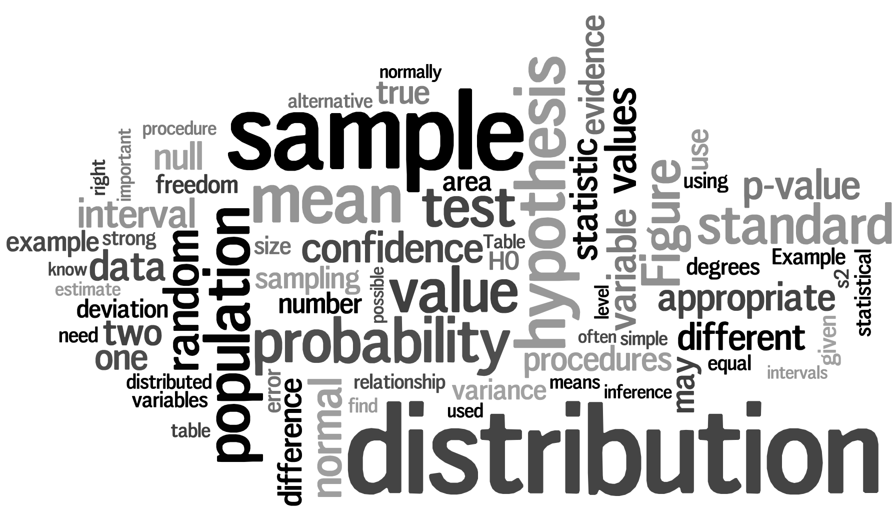
Introductory Statistics Explained
Exercises

Edition 1.09CC



See http://www.jbstatistics.com/ for the complete set of supporting videos and other supporting material.

The complete set of videos is also available at:
https://www.youtube.com/user/jbstatistics/videos

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. Succinctly state the difference between a parameter and a statistic.

2. Researchers are interested in estimating the average weight of three-year-old walleye in a lake. A netting program yielded 52 three-year-old walleye, with an average weight of 273 grams.

   (a) What is the population of interest?
   (b) What is the sample?
   (c) Is the value 273 a parameter or a statistic?
   (d) What is the parameter of interest in this scenario?
   (e) Is it possible to know the value of the parameter of interest in this scenario?

2.3 Types of Sampling

2.3.1 Simple Random Sampling

3. An importer of used medical equipment wants to estimate the proportion of defective pacemakers in a shipment of 200 pacemakers. They open the shipping container, remove the top 8 pacemakers, and find that 0% of of these 8 pacemakers are defective. Unknown to the importer, 10% of the pacemakers in the entire shipment are defective.

   (a) What is the population of interest?
   (b) What is the sample?
   (c) What value given above is a *statistic*?
   (d) What value given above is a *parameter*?
   (e) Would this sample above constitute a *simple random sample*?

4. Suppose a professor wishes to estimate the proportion of female students in a class of 50 students. They could find out the true proportion without too much difficulty, but they decide instead to draw a sample of 10 students. Each student's ID number is put on a marble. The 50 marbles are put into a box, and the box is shaken vigorously. Suppose after the vigorous shaking that the marbles can be considered to be randomly distributed in the box. Without looking at the marbles, the professor reaches into this box and pulls out 10 marbles without replacement. The 10 students with these ID numbers are a sample from the population of 50 students. Which, if any, of the following statements are true?

   (a) The sample can be considered to be a simple random sample from the population.
   (b) Each possible sample of size 10 has the same chance of being selected.
   (c) The 50 students represent the population.
   (d) The probability any individual in the population will be in the sample is 0.20.

2.3.2 Other Types of Random Sampling

5. A professor is interested in learning more about her students, and she decides that she is going to draw a sample of 10 students from her population of 50. She knows she should be drawing some sort of random sample, but doesn't quite know how to go about it. She considers the following two sampling designs.

   I. Put all the names of students on small pieces of paper, put them into a hat, and mix them up such that the names are randomly distributed. Then she

will pick 10 names without replacement.

II. To ensure equal representation of males and females, she will have separate hats for males and females. In the first hat she will mix up the names of the male students, then draw 5 names without replacement. She will carry out the same procedure with the females, drawing 5 female names. She will then pool these names together into a group of 10.

Which of the following statements are true?

(a) Sampling method I results a simple random sample.
(b) Sampling method II results in a stratified random sample.
(c) Both sampling designs result in very biased samples.

2.4 Experiments and Observational Studies

6. A doctor is investigating the relative effectiveness of two different procedures that may prolong life in patients with advanced heart disease. Twenty volunteers are randomly assigned to the two procedures, 10 to procedure A and 10 to procedure B. The time until death is recorded.

   (a) Is this an experiment or an observational study?
   (b) What are the experimental units?
   (c) What are the treatments?
   (d) What is the response variable?
   (e) If it is found that those patients from procedure B tended to live much longer on average, is it reasonable to conclude that procedure B caused this increase?

7. Researchers are interested in the yield of four types of corn. There are 20 plots of land available, and the researchers randomly assign the types of corn to the 20 plots of land (5 plots for each type of corn). After the growing season, the yield (kg) is measured.

   (a) Is this an observational study or an experiment?
   (b) What are the treatments?
   (c) What is the response variable?
   (d) What are the experimental units?

8. Suppose a long-term observational study finds that children who watch an average of more than 50 hours of television a week are much more likely to be arrested by age 18 than those children who watch less television. Based on this study, is it

reasonable to conclude that increased television viewing causes an increase in the likelihood of arrest?

2.5 Chapter Exercises

9. A study[1] investigated fast food consumption among students on a college campus in the southern United States. The researchers recruited students by posting flyers on campus, offering to give participants small giveaways and a chance to win a $100 gift card from the campus bookstore. The 152 student participants completed several surveys, and one of the surveys involved fast food consumption. In the analysis of the results of one of the surveys, the researchers found that male students tended to spend more per month on fast food than female students did.

   (a) What is the population of interest?
   (b) Is the sample of students a simple random sample?
   (c) What types of bias might be present in a study like this?
   (d) Would it be possible to conduct a simple random sample of students at this university to investigate fast food consumption?

10. A study[2] investigated physical characteristics of two species of lizard (*Phrynocephalus frontalis* and *P. versicolor*) found in a region of Inner Mongolia. The researchers were interested mainly in the differences between males and females within each species. Researchers captured these lizards by hand or by noose, and measured various physical characteristics.

    (a) What are the populations of interest?
    (b) Did the researchers use a simple random sampling design?
    (c) What types of bias might be present in this study?

11. A study[3] investigated a possible effect of a vitamin C supplement on the frequency and duration of colds. One thousand volunteers were randomly assigned to two groups. One group received a vitamin C supplement of 1g per day, the other group received a placebo. The participants then kept track of variables related to cold frequency and severity for the duration of the study (approximately 2 months).

    (a) This experiment was conducted as a *double-blind* experiment (neither the

[1] Heidal et al. (2012). Cost and calorie analysis of fast food consumption in college students. *Food and Nutrition Sciences*, 3:942–946

[2] Qu et al. (2011). Sexual dimorphism and female reproduction in two sympatric toad-headed lizards, *Phrynocephalus frontalis* and *P. versicolor* (Agamidae). *Animal Biology*, 61:139–151

[3] Anderson, T., Reid, D., and Beaton, G. (1972). Vitamin C and the common cold: a double-blind trial. *Canadian Medical Association Journal*, 107:503–508

study participant nor the researchers involved knew whether the participant was receiving the vitamin C or the placebo, until after the study was completed and the data was summarized). Why would the researchers set up the experiment in this way?

(b) The study began with 1000 participants, but 182 people dropped out before the end of the study. The data for the individuals that dropped out was thrown out and the analysis was based on the remaining 818 participants. How might this bias the results?

12. Researchers investigated a possible effect of exercise on tumour growth in rats. Thirty rats were injected with cancerous cells that would develop into a tumour. These rats were then randomly assigned to one of two groups (15 to each group). Rats in one of the groups were forced to run on a treadmill for 30 minutes per day. Rats in the other group had no forced exercise. After one month, the rats were sacrificed and the weight of the tumour was measured.

    (a) What is the response variable?
    (b) What is the explanatory variable?
    (c) On one of the days, a rat caught his foot on the back of the treadmill and was badly injured. The rat had to be sacrificed and omitted from the experiment. How might this bias the results? Does this render the results of the study invalid?

13. Consider the following list of Central American countries:

    - Belize
    - Costa Rica
    - El Salvador
    - Guatemala
    - Honduras
    - Nicaragua
    - Panama

    (a) Number the countries from 1–7, then pick a simple random sample of size 3 from this group using the following random numbers:
    98797 18334 57628 33215 32849.
    What countries are selected? (Hint: We would ignore the first 2 numbers in the list (9,8), since no country is labelled with a 9 or an 8, and the first country selected would be Panama (7).)
    (b) Suppose instead of the above, we randomly pick a letter from A-Z using a random number generator, then include the country in our sample if the country's name begins with that letter. If we continued in this fashion until 3 countries were selected, would that sample represent a *simple random sample* from the

population?

14. None of the sampling methods in this question are perfect, but which one would yield the most representative sample from its population? (This question requires a little thought, and the correct answer is subject to debate.)

    (a) A student wants to know the class average on a midterm exam she recently wrote. At the lecture following the midterm, Friday at 8:30, she arrives to class early and asks ten students in the lecture room what their mark was on the test. She calculates her estimate of the class average based on what these 10 students say they scored on the test.
    (b) A student wants to investigate the average yearly income of the parents of university students. While in line to obtain his student loan, he asks 10 people in line with him what their parents' yearly income is, and takes an average of the responses.
    (c) A teaching assistant wants to know what proportion of students in a class are male. This particular teaching assistant is a little lazy, and simply looks at the first 10 students on a class list that is sorted alphabetically by last name, and finds the proportion of these students that are male.

15. Researchers are interested in estimating the proportion of full-time female graduate students at a university that have a job outside of the university. Surveys are sent out to all full-time female graduate students, and 67 surveys are completed and returned. Of these 67 students, 19 have jobs outside of the university.

    (a) What is the population of interest?
    (b) What is the sample?
    (c) What is the parameter of interest?
    (d) What is the statistic that estimates this parameter?
    (e) Is this an experiment?
    (f) Is this a simple random sample?
    (g) Will this sample be representative of the population as a whole?

16. Suppose that a person is interested in estimating the average weight of black bear gall bladders in Ontario. They have a circle of hunting friends, and they find that their next 3 kills, the bears have an average gall bladder weight of 22 grams. Suppose that, in reality, the average weight of Ontario black bear gall bladders is 24 grams. (Please don't read into this question as support for bear hunting or the sale of parts, because it's not.)

    (a) What is the population of interest?
    (b) Is 22 grams the value of a parameter or a statistic?
    (c) Is 24 grams the value of a parameter or a statistic?

(d) The sampling method could best be described as:
  i. Simple random sampling.
  ii. Stratified random sampling.
  iii. Voluntary response sampling.
  iv. Something else.

17. Suppose a botanist has grown 11 sunflowers and wishes to randomly select 2 of them for further study. She orders the 11 sunflowers from largest to smallest, and numbers them from 02 to 12 (in that order). She doesn't have access to a computer or a random number table, but she does have a pair of dice available. She rolls the two dice, gets the sum of the numbers on the two faces, and uses that number to select the first sunflower for her sample. She rolls again to randomly select a second sunflower. Note that the chances of rolling the various possible sums are different (the probability of rolling numbers that sum to 2 is 1/36, whereas the probability of rolling numbers that sum to 7 is 6/36).

    Do the two sunflowers that the botanist selected represent a simple random sample from the population of 11 sunflowers?

18. Suppose a pro-gun organization in the U.S. sends a survey to 1000 of its members, asking "Do you believe that all illegal immigrants should be deported back to their country of origin?". 912 members return the survey, 59% responding "Yes". This survey is then used as evidence that 59% of Americans believe that illegal immigrants should be deported. (The survey results given in this question are fictitious and are not meant to be accurate.)

    Which of the following statements are true? (check all that are true—there may be more than one correct statement)

    (a) Since only 912 of the 1000 people who were sent the survey returned it, the survey gives absolutely no evidence about the feelings of members of this organization toward illegal immigrants.
    (b) The biggest problem with this survey is that it is a mail-in survey, which can cause a delay of several weeks.
    (c) The sample size is far too small to be meaningful.
    (d) The sample is not representative of the population of interest.

19. A principal wants to estimate the average IQ of third graders in her school. The first 10 third graders on the alphabetical list are selected, and their average score on an IQ test is found to be 97.2.

    (a) What is the population of interest?
    (b) Is the number 97.2 a parameter, or a statistic?
    (c) Is the principal's sample a *simple random sample*?

20. Suppose a very conservative media outlet in the U.S. asks a survey question of its viewership, "Do you approve or disapprove of the way the president is handling his job as President of the United States?". This poll has 1500 respondents, 72% of this sample disapproving of the way the president is handling his job as president. Suppose this statement is then used to imply that 72% of Americans disapprove of the way the president is handling his job.

    Which of the following statements are true?

    (a) The 1500 respondents represent a simple random sample.
    (b) The wording of the question is very misleading and would inevitably result in a biased sample.
    (c) The number 72% is a parameter.
    (d) Since only 1500 people were sampled from a population of over 300 million, the sample size is far too small to be meaningful.

21. A study investigated the effect of a regular dose of aspirin on the risk of heart attack. Physicians randomly assigned approximately 11,000 people to a group that took an aspirin every second day, and 11,000 to a control group that took a placebo instead. After several years the subjects in the aspirin group were found to have a lower risk of heart attack.

    (a) Is this an experiment or an observational study?
    (b) What is the response variable?
    (c) What is the explanatory variable?
    (d) Is it necessary to have exactly half the subjects in one group and half in the other?
    (e) True or False: The heart attack variable is confounded with the aspirin variable.

22. Consider again the information in Question 21. Suppose that instead of carrying out an experiment, the researchers sent out a survey to physicians, asking them if they had taken aspirin regularly over the past 5 years, and if they had had a heart attack in that time.

    (a) If the data from this survey established a link between aspirin use and a lower rate of heart attack, would the conclusions be as strong as those found in Question 21?
    (b) What sort of biases are possible in this survey?

23. Researchers wish to investigate the effect of negative reinforcement (a mild electric shock) on the ability of people to concentrate. Twenty people volunteer for this investigation. Ten are randomly assigned to a shock group, and ten are assigned to a control group. Each individual plays a memory game, involving 100 tiles laid face

down on a table. The 100 tiles each have a picture on the face-down side. There are 50 pairs of identical pictures, and the individual must attempt to find matching pairs. Individuals assigned to the shock group get an electric shock whenever their attempted match is wrong. Individuals in the control group do not get a shock. Each individual is allowed 100 attempts at finding matching pairs of tiles, and the number of matching pairs they find is recorded.

Which of the following statements are true?

(a) There are 20 treatment groups.
(b) This is an observational study.
(c) Even if there is strong evidence of a difference between the two groups, there will not be evidence of a causal effect.
(d) The response variable is the voltage of the electric shock.

24. A company is concerned that there are too many injuries on one of their assembly lines. They decide that they may need to make certain workers wear a type of harness to prevent falls and injuries. They are concerned that the harness may cause the workers to work more slowly and have longer task completion times. Fifteen workers volunteer for a study to investigate this. Eight of these workers are randomly assigned to the harness-wearing group, and the remaining seven workers will not wear a harness. The workers are observed for two weeks and their task completion times are recorded.

(a) What is the response variable?
(b) What is the explanatory variable?
(c) Is this an experiment or an observational study?
(d) Can this type of study help determine if there is a causal link between wearing the harness and slower times?
(e) Why did the company randomly assign the workers to the groups, and did not allow the workers to pick which group they wanted to be in?

25. You want to perform an experiment to investigate the effect of different stimuli on the time it takes for rats to complete a maze. You decide to have 3 groups of rats. Group 1 is a control group, rats in Group 2 will receive a piece of cheese if they complete the maze within a certain time, and rats in Group 3 will get a mild (but painful) electric shock every 3 seconds until they complete the maze. You have 9 rats available for the experiment, and you (wisely) decide to randomly assign 3 rats to each group. You have named the rats:

Pete, Tom, Jerry, Alphonse, Renaldo, Bo, Little Pete, Huey, Eddie.

Number the rats from 1–9 using a single digit. Then randomly assign the rats to the groups by using the following random numbers.

Random Number Line: 33696 91544 76248

(The first 3 chosen go to the control group, the next 3 to the cheese group, and the remainder are the unlucky ones in the electric shock group.)

(a) Which rats go in which group?
(b) Little Pete is your favourite rat, and you do not want to see him shocked. You decided to put him in the cheese group, and randomize the others. Why is this not acceptable from a statistical point of view?

26. Suppose a new yoga centre conducts a survey of 1000 women. They find that those who practice yoga regularly are much healthier than those that do not. They use this finding in their advertising, stating that their survey shows that "regularly practicing yoga can improve women's health". What, if anything, is wrong with that statement?

27. Which one of the following is the most reasonable definition of a lurking variable?

(a) An unmeasured variable that is related to both the response and explanatory variables.
(b) An unmeasured variable that is related to the response variable, but has no relationship with the explanatory variable.
(c) A measured variable that affects the response, but is not related to the explanatory variable.
(d) A variable that has a strong effect on the mean of a data set, but does not affect the median.
(e) An explanatory variable that causes changes in the response variable. Experiments can give strong evidence of a causal relationship between variables.

28. Researchers are interested in assessing the effect of four different drugs on patients with high blood pressure. Two hundred people with high blood pressure volunteer for an experiment. The 200 people are randomly assigned to the different drugs, 50 to each drug. After 12 weeks on the drug, the reduction in blood pressure is recorded.

(a) What is the response variable?
(b) What is the explanatory variable?
(c) Is this an experiment, or an observational study?
(d) Can this type of study give strong evidence of a causal link between the drug that is taken and reduced blood pressure?

29. An insurance company lets new policy purchasers choose whether they wish to purchase "accident forgiveness", which allows the driver to avoid a premium increase in the event of an at-fault accident. Out of 200 new purchasers, 40 choose to

purchase accident forgiveness. It is found that after one year, 8 of the 40 who purchased accident forgiveness got into an at-fault accident, whereas only 4 of the 160 that did not purchase it got into an at-fault accident. Which of the following statements are true? There may be more than one true statement; check all that are true.

(a) This is an observational study, not an experiment.
(b) There is very strong evidence that purchasing accident forgiveness causes more drivers to get into an at-fault accident.
(c) This study shows that there is definitely no relationship between purchasing accident forgiveness and getting into an accident.

30. Dental researchers are investigating the strength of two new bonding agents, and comparing their bond strength to that of the commonly used (standard) bonding agent. They run an experiment in which the 3 compounds are tested on extracted teeth. There are 60 extracted teeth available for the experiment. The 60 teeth are randomly assigned to the 3 bonding compounds, 20 to each compound. The researchers attach a dental appliance to each tooth, and the bond strength is recorded.

    (a) What are the treatments?
    (b) What is the explanatory variable?
    (c) Is this an observational study, or an experiment?
    (d) Can this type of study help establish if a there is a causal link between the bonding agent and the strength of the bond?

31. Consider the following scenarios. In each scenario determine whether it would be best investigated with an experiment, a survey, or an observational study that is not a survey.

    (a) You are conducting a study investigating drivers in fatal single-vehicle accidents. You want to determine whether male drivers are more likely than female drivers to have alcohol in their system.
    (b) You want to determine whether new male MBA graduates have a higher salary on average than new female MBA graduates.
    (c) You want to determine which of two commonly used headache medications results in the fastest pain relief.
    (d) You want to determine if there is a relationship between cocaine use and marital infidelity.
    (e) You want to investigate whether men who commit suicide tend to use a firearm more frequently than women who commit suicide.

32. A veterinarian is developing a new experimental surgical technique for dogs that have a certain type of cancer. The surgery is very risky, and many of the dogs die soon after surgery. But the veterinarian believes that most dogs would live longer

after the experimental surgery than if they had the standard surgery. She offers dog owners a choice of the two methods. She finds that those dogs that undergo the new experimental surgery live a shorter time on average than those that undergo the standard surgery.

Disheartened, she decides to carry out an experiment. Dogs in the experiment are randomly assigned to one of the two surgical methods. In the experiment, it is found that dogs that undergo the experimental surgery live much longer on average than dogs that undergo the standard surgery.

(a) If you have a dog with bone cancer, and your sole objective is to maximize their length of life after surgery, which surgery method should you choose?
   i. The new experimental surgery.
   ii. The standard surgery.

(b) Which one of the following is the most important *lurking variable* that would affect the interpretation of the results of the observational study?
   i. The medical condition of the dog.
   ii. The size of the dog.
   iii. The income of the dog owners.
   iv. The IQ of the veterinarian.
   v. The breed of the dog.

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 Finnish study[1] investigated a possible association between the gender of convicted murderers and their relationship to the victim. A random selection of 91 female murderers and a random selection of 91 male murderers were obtained, and the results are illustrated in Figure 3.1.

   (a) For male murderers, summarize the distribution of the relationship between the murderer and the victim.
   (b) Describe the observed differences in the distributions of the relationship to the victim between male and female murderers.
   (c) Can we be certain that the observed differences in the samples of male and female murderers reflect the true differences in the population distributions?
   (d) Sketch a pie chart to illustrate the distribution of the relationship to the victim for male murderers. (Sketch a rough plot – it does not need to be very

[1] Häkkänen et al. (2009). Gender differences in Finnish homicide offence characteristics. *Forensic Science International*, 186:75–80.

Gender	Male	Female	Total
Acquaintance	61	37	98
Partner	22	32	54
Family Member	4	18	22
Stranger	4	4	8
Total	91	91	182

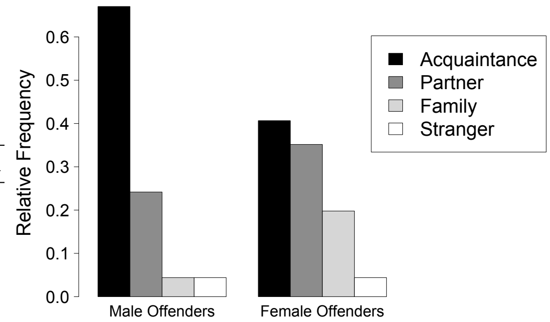


Figure 3.1: Gender of murderer and their relationship to the victim in Finnish murders.

accurate.)

3.2.2 Plots for Quantitative Variables

3.3 Numerical Measures

3.3.1 Summation Notation

2. Suppose we have the following sample data set: 8, 14, 22, −5

   (a) What is the value of x_3?

   (b) What is $\sum_{i=1}^{3} x_i$?

   (c) What is $\sum x_i$?

   (d) What is $\sum x_i^2$?

3.3.2 Measures of Central Tendency

3.3.2.1 Mean, Median, and Mode

3. Suppose we have a sample of 5 observations: 1, 5, 2, −3, 987.

   (a) What is the mean?

   (b) What is the median?

   (c) What is the mode?

(d) If the extreme value is removed, what is the mean?
(e) If the extreme value is removed, what is the median?

4. A random sample of 4 Canadian male newborns revealed the following birth weights, in grams:

$$2870, 2620, 3120, 3620$$

(a) What is the mean birth weight?
(b) What is the median birth weight?
(c) What are the units of the mean?
(d) What are the units of the median?
(e) In this situation, which is the more appropriate measure of central tendency, the mean or the median?

3.3.2.2 Other Measures of Central Tendency

5. What would be an advantage of using the trimmed mean instead of the untrimmed mean? What would be a disadvantage?

3.3.3 Measures of Variability

6. Consider the following sample of 4 observations: 18, 8, 3, 17.

(a) What are the deviations?
(b) What is the sum of the deviations?

7. A random sample of 4 Canadian male newborns revealed the following birth weights, in grams:

$$2870, 2620, 3120, 3620$$

(a) What is the range?
(b) What is the mean absolute deviation?
(c) What is the variance?
(d) What is the standard deviation?
(e) What are the units of the variance?
(f) What are the units of the standard deviation?

8. Create a 4 number sample data set, where all numbers lie between 0 and 500 (inclusive, and repeats are allowed), such that the standard deviation is as large as possible.

(a) What is your data set?
(b) What is the standard deviation of your data set?

9. Create a 4 number sample data set, where all numbers lie between 0 and 500 (inclusive, and repeats are allowed), such that the standard deviation is as small as possible.

   (a) What is your data set?
   (b) What is the standard deviation of your data set?

10. Which of the following statements are true? There may be more than one correct statement; check all that are true.

   (a) The standard deviation can be greater than the variance.
   (b) The standard deviation can be negative.
   (c) The standard deviation can be less than the mean.
   (d) The standard deviation can be less than the third quartile.
   (e) The standard deviation is always less than the average distance from the mean.

3.3.3.1 Interpreting the standard deviation

11. Figure 3.2 illustrates scores on a difficult statistics test. The scores have a mean of 22.4 and a standard deviation of 7.2. (The maximum possible score on the test was 40.)

   (a) Would the empirical rule apply to this data? Why or why not?
   (b) What would the empirical rule tell us about the proportion of observations that are within 7.2 of 22.4?
   (c) What would the empirical rule tell us about the proportion of observations that are within 14.4 of 22.4?
   (d) What would the empirical rule tell us about the proportion of observations that are within 21.6 of 22.4?

12. Figure 3.2 illustrates scores on a difficult statistics test. The scores have a mean of 22.4 and a standard deviation of 7.2.

   (a) Would Chebyshev's theorem apply to this data? Why or why not?
   (b) What would Chebyshev's theorem tell us about the proportion of observations that are within 7.2 of 22.4?
   (c) What would Chebyshev's theorem tell us about the proportion of observations that are within 14.4 of 22.4?
   (d) What would Chebyshev's theorem tell us about the proportion of observations that are within 21.6 of 22.4?

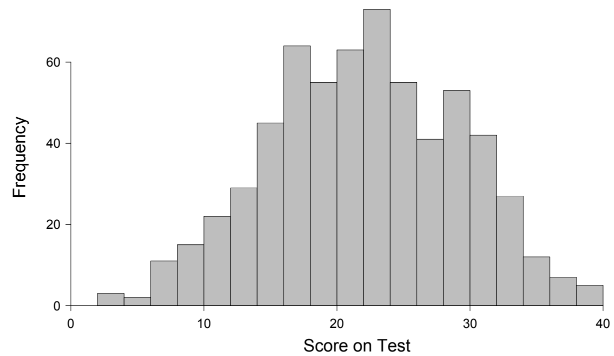


Figure 3.2: Scores on a hard test ($\bar{x} = 22.4$, $s = 7.2$).

13. Consider the histogram given in Figure 3.3.

    (a) Would the empirical rule apply to this data?
    (b) Would Chebyshev's theorem apply to this data?

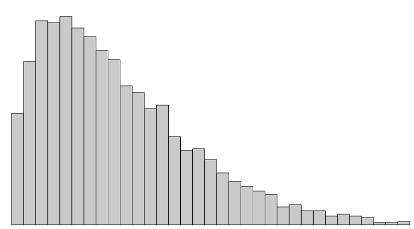


Figure 3.3: A distribution that is skewed to the right.

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

14. Why do we divide by $n-1$ in the sample variance formula?

15. Suppose we have a sample of size $n = 87$, and the population mean and variance are unknown. How many degrees of freedom are there for estimating the variance?

3.3.4 Measures of Relative Standing

3.3.4.1 Z-scores

16. A random sample of 4 Canadian male newborns revealed the following birth weights, in grams:

$$2870, 2620, 3120, 3620$$

In Exercise 7, we found that for these four observations: $\bar{x} = 3057.5$ and $s = 426.9563$.

(a) What are the 4 z-scores?
(b) What is the mean of the 4 z-scores?
(c) What is the standard deviation of the 4 z-scores?
(d) If a newborn male baby had a z-score of 4.6, what does that tell us about the baby's weight?
(e) If a newborn male baby had a z-score of -0.4, what does that tell us about the baby's weight?

17. Todd has always had a dream of becoming a medical doctor. After doing well in an introductory statistics course, Todd decides to write the MCAT. His score on the test corresponded to a z-score of 3.0. Suppose that scores on the test have a distribution that is mound-shaped (approximately normal). Which of the following statements are true?

(a) The z-score is a unitless quantity.
(b) Todd's score was 3 standard deviations greater than the mean score.
(c) Todd scored worse than approximately 1/3 of the test writers.
(d) Todd's score was better than average.

3.3.4.2 Percentiles

18. A sample of 8 boxes of Kellogg's All Bran was collected at a grocery store. The boxes had a nominal weight of 675 grams. The weight (in grams) of the cereal in each box was recorded, with the following results:

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

(The weights were recorded after discarding the box and the bag, so they represent the weight of only the cereal.)

(a) Use the method outlined in the text to calculate the 80th percentile of the weights.
(b) Use the method outlined in the text to calculate the 25th percentile of the weights.

19. The 90th percentile of heights of adult females in the United States is closest to which one of the following?

(a) 90 cm.
(b) 122 cm.
(c) 143 cm.
(d) 171 cm.
(e) 200 cm.

3.4 Boxplots

20. Consider the boxplots in Figure 3.4, representing 3 different samples.

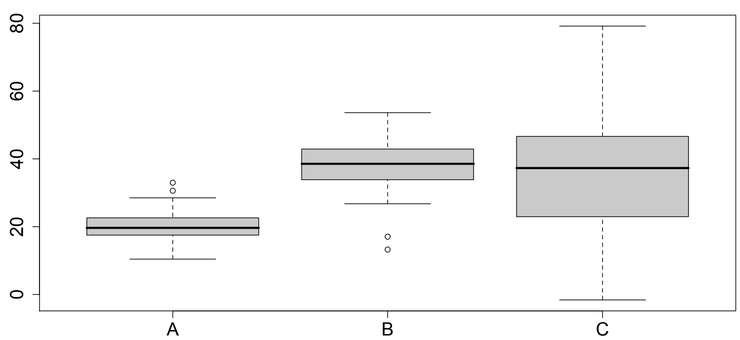


Figure 3.4: 3 boxplots.

(a) What is the median of sample C?
(b) What is the range (Maximum – Minimum) of sample C?
(c) How many outliers are there in the entire plot (all samples combined).
(d) What is the 25th percentile of sample C?

21. Qu et al. (2011) investigated physical characteristics of the lizard *Phrynocephalus frontalis*. In one part of the study, the researchers compared the tail lengths of males and females of this species. Figure 3.5 illustrates the distributions of tail length for 44 female and 22 male lizards that were captured in the wild.

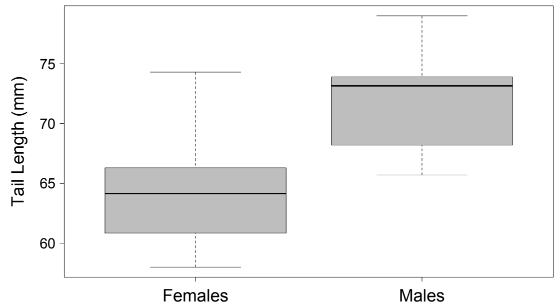


Figure 3.5: Boxplots of tail lengths of 44 female and 22 male lizards of the species *P. frontalis*.

(a) What is the 75th percentile of tail lengths for the sample of male lizards?
(b) What is the 75th percentile of tail lengths for the sample of female lizards?
(c) Summarize the main differences and similarities between males and females for these samples of tail lengths.

3.5 Linear Transformations

22. A certain sample has a mean of 100, a median of 90, and a standard deviation of 20. Suppose that each observation has 5 added to it, and the result is multiplied by 20.

    (a) What is the mean of the new values?
    (b) What is the variance?
    (c) What is the standard deviation?
    (d) What is the median?

23. Suppose we have a sample of 7 values, and we calculate the usual summary statistics. If we add 5,000 to each of these 7 sample values, then recalculate the summary statistics, which of the following quantities would change?

    mean, median, Q_3, IQR, variance, standard deviation

3.6 Chapter Exercises

3.6.1 Basic Calculations

24. Consider the following sample of 4 numbers: $52, -2, 8, 107$.
    (a) What is the mean?
    (b) What is the median?
    (c) What is the variance?
    (d) What is the standard deviation?
    (e) What is the range?

25. Consider the following sample of 7 observations:

$$3.1,\ 8.2,\ 9.6,\ -1.7,\ 8.4,\ 8.8,\ 21.1$$

    (a) What is the mean?
    (b) What is the median?
    (c) What is the third quartile (Q_3)?
    (d) What is the interquartile range?
    (e) What is the variance?
    (f) What is the standard deviation?
    (g) Are any of these observations outliers according to the 1.5*IQR rule?

26. Consider the following sample data set: $32, 36, 1, 4, 89, 18$.
    (a) What is the mean?
    (b) What is the median?
    (c) What is the value of Q_1?
    (d) What is the value of the interquartile range?
    (e) What is the variance?
    (f) What is the standard deviation?

27. Consider the following sample data set: $-4, 6, -12, 14, 742$
    (a) What is the mean of the full data set?
    (b) What is the median of the full data set?
    (c) What is the standard deviation of the full data set?
    (d) What is the variance of the full data set?
    (e) If the outlier is discarded, what is the mean?
    (f) If the outlier is discarded, what is the median?
    (g) If the outlier is discarded, what is the standard deviation?

28. Consider the following computer output, illustrating a stemplot.

```
Decimal point is at the colon

  -1 : 7110
  -0 : 76632
   0 : 03335
   1 : 01389
   2 : 4
```

(a) What is the mean?
(b) What is the median?
(c) What is the variance?
(d) What is the standard deviation?
(e) What is the range?

29. Consider the following split-stem stemplot: (The decimal is at the colon. You should be able to recognize that the largest observation is 5.8.)

```
  0 : 000000000000000000111111222222233334444444
  0 : 557777888
  1 : 001111224444
  1 : 5555778999
  2 : 011223
  2 : 55566689
  3 : 0
  3 : 5678
  4 :
  4 : 7789
  5 :
  5 : 668
```

(a) Is the mean greater than the median?
(b) Does the distribution show any skewness? If so, in what direction is it skewed?
(c) What is the value of the difference between the fifth and first values in the five-number summary?

30. Consider the following stemplot:

```
-4 : 975
-3 : 99410
-2 :
-1 : 6
-0 :
 0 :
 1 : 1
 2 : 13
 3 : 249
```

(The decimal place is at the colon.)

(a) What is the mean?
(b) What is the median?
(c) If the decimal place was shifted one place to the left for all the observations (e.g. the smallest observation was -0.49), which of the following statistics would change?

Mean, median, Q_1, Q_3, IQR, standard deviation, variance

3.6.2 Concepts

31. Consider the bimodal frequency histogram for a variable x, illustrated in Figure 3.6.

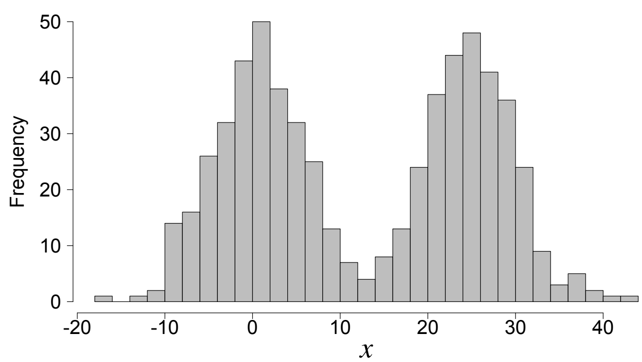


Figure 3.6: A bimodal frequency histogram.

(a) What is the approximate value of the median?
(b) What is the approximate value of Q_1?
(c) What is the approximate value of Q_3?
(d) Are there any outliers?

(e) True or false: The mean and median would be close in value.
(f) One of the boxplots in Figure 3.7 represents this data. Which one?

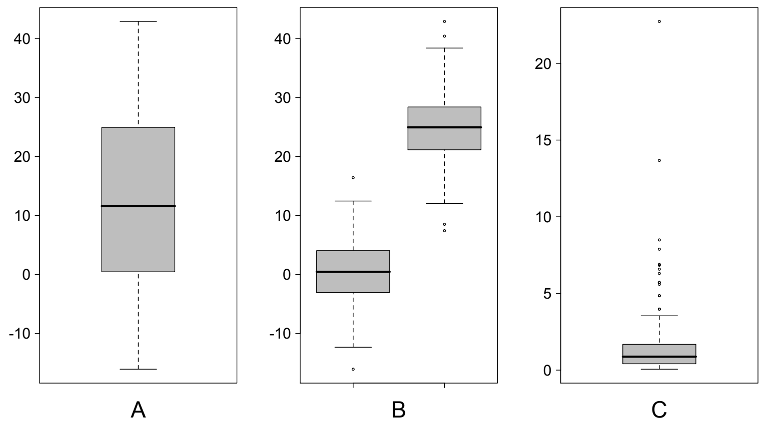


Figure 3.7: Which boxplot corresponds to the histogram?

32. Suppose we have a sample of 5 observations. We forget the data values, but we know the first 4 deviations from the mean $(x_i - \bar{x})$ are:

$$-2.3, 3.7, 1.2, -0.7$$

(a) What must the value of the fifth deviation be?
(b) What was the sample mean?
(c) What is the standard deviation?

33. If the smallest 4 values in a sample with $n = 5$ are equal, and the range is equal to 10, what is the value of the variance?

34. Consider the frequency histogram shown in Figure 3.8. Which of the following statements are true?

(a) The median is less than 4.
(b) The mean is less than the median.
(c) The IQR is greater than 5.
(d) All the numbers in the five number summary are positive.
(e) This distribution is approximately symmetric.
(f) The value of $Q_1 -$Median would have the same absolute value as $Q_3 -$Median.
(g) There are no obvious outliers.

35. Consider the boxplots in Figure 3.9.

(a) Which one of the boxplots corresponds to the histogram in Figure 3.8?

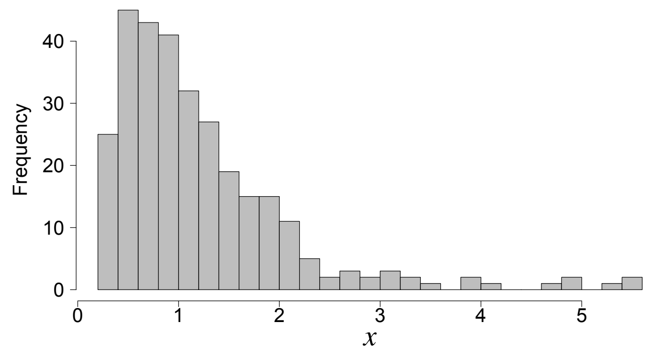


Figure 3.8: A frequency histogram.

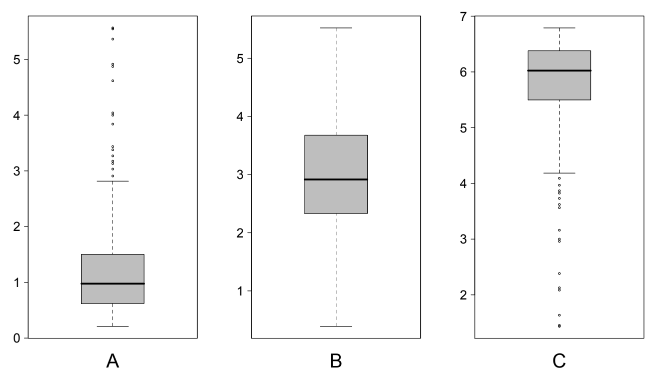


Figure 3.9: Which boxplot represents the histogram data?

(b) Which boxplot has the largest median?
(c) Which boxplot has the largest interquartile range?
(d) Which boxplot contains the largest number of outliers?

36. Suppose we have a sample of 12 observations, and we calculate the summary statistics for these values.

(a) If we add 100,000 to the largest observation, which of the following statistics would change?
Mean, median, variance, standard deviation, interquartile range
(b) If we add 100,000 to the largest observation, and subtract 100,000 from the smallest observation, which of the following statistics would change?
Mean, median, variance, standard deviation, interquartile range

37. Suppose there is a data set of 100 observations that has a standard deviation equal to 0. Which of the following statements are true?

(a) All the numbers in the 5 number summary must be equal.

(b) There are no negative deviations.
(c) The variance and standard deviation are equal.
(d) The mean and median are equal.
(e) If a single observation from this data set had a positive constant added to it, then the IQR, minimum, and median would not change.

38. Suppose we have a sample data set with 4 observations, where $\bar{x} = 50$ and $s = 0$. Suppose we add one more observation to this data set, and we recalculate the mean and standard deviation. If the added observation equals 50.0, what will the standard deviation of the 5 number data set be?

39. Which of the following statements are true?

    (a) The standard deviation is always greater than the median.
    (b) The variance is always greater than the standard deviation.
    (c) If a distribution is skewed to the left, then the mean will always be less than the standard deviation.
    (d) If all values in a sample are within 2.0 of each other, then the standard deviation will be less than 2.0.
    (e) The median can equal the third quartile (Q_3).

40. Which of the following statements are true?

    (a) If the median of a distribution equals the minimum, then the distribution can't be skewed to the left.
    (b) If a variable x is measured in seconds, then the units of the variance are seconds^2.
    (c) The 12.5$^{\text{th}}$ percentile is exactly half the value of Q_1.
    (d) All of the values in the five-number summary are always positive.
    (e) Large outliers have little effect on the standard deviation, but can have a strong effect on the variance.

41. Which of the following statements are true?

    (a) For right-skewed distributions, the median is greater than the mean.
    (b) For symmetric distributions, the standard deviation and mean are equal.
    (c) For mound-shaped distributions, approximately 50% of observations lie within 5 standard deviations of the mean.
    (d) For any distribution, the interquartile range is equal to 50.
    (e) All observations that are less than Q_1 would have negative z-scores. (Careful!)
    (f) If the range of a data set is equal to 0, then the variance and IQR will also equal 0.

42. Consider the following sample data:

$$-2, -5, 12, 17, 15, 15, 15, 15, 15, 15, ?, ?$$

(The question marks represent two missing values.)

(a) What is the value of the median for the entire sample (including the missing values)? If it is impossible to determine, say so.

(b) What is the value of the mean for the entire sample (including the missing values)? If it is impossible to determine, say so.

(c) What is the value of the variance for the entire sample (including the missing values)? If it is impossible to determine, say so.

43. A certain data set has a mean of 15 and a standard deviation of 5. After a linear transformation, the new standard deviation was found to equal 1, but the mean was unchanged.

(a) What possible linear transformations result in these new values (there are two possibilities).

(b) Suppose the appropriate transformation is the one where b is positive. If a value in the old data set is 10, what is its new value?

44. In a sample of size $n = 3$, the standard deviation is equal to 100. Two of the observations are 152 and 252. The third observation is one of two possible values. What are these two possible values?

3.6.3 Applications

45. The histogram in Figure 3.10 represents final grades for students in an introductory statistics course. (This is real data from one of your author's courses.)

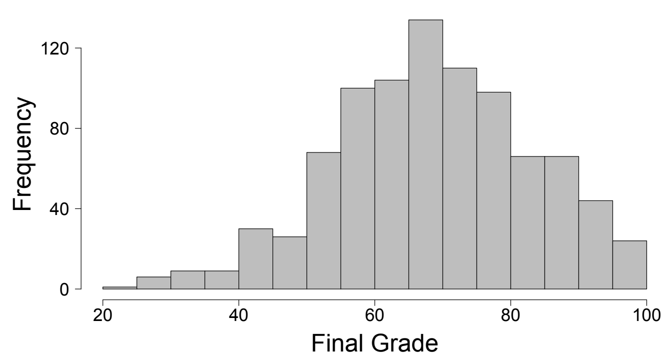


Figure 3.10: Final grades in an introductory statistics course.

(a) What is the approximate value of the median?
(b) What is the approximate value of Q_1?
(c) What is the approximate value of the interquartile range?
(d) Would the mean be greater or less than the median?
(e) What is the approximate value of the standard deviation?

46. Figure 3.11 is a histogram of grades on a major assignment in a large introductory statistics course. (This is real data from one of your author's courses.) Of the 206 students in the course, twelve students did not hand in the assignment and received a grade of 0. The average grade for all students was 18.8 out of 25.

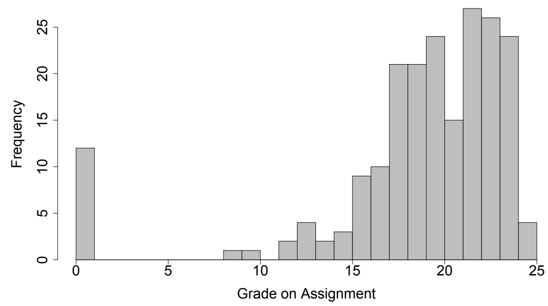


Figure 3.11: Grades on a major assignment.

(a) Describe the distribution.
(b) Reporting the mean grade for all students might be slightly misleading in this setting. Suggest a better numerical summary of this data.

47. The insect *Megamelus scutellaris* can feed on the sap of the water hyacinth, and it has been used as a method of biocontrol for this invasive plant species. *M. scutellaris* mates on the water hyacinth, and females create oviposition scars on the plant, laying one or more eggs in each scar. Sosa et al. (2005) investigated several aspects of the reproductive cycle of this insect. In one aspect of this study, the number of eggs laid per scar was measured. The observed number of eggs per scar for 359 scars is given in Table 3.1.

1 egg/scar	2 eggs/scar	3 eggs/scar	4 eggs/scar
142	194	9	14

Table 3.1: Observed number of *M. scutellaris* eggs per scar.

(a) What is the mean number of eggs per scar?
(b) What is the median number of eggs per scar?
(c) What is the standard deviation of the number of eggs per scar?

48. As part of an investigation by Fink et al. (2007), 32 male student volunteers at a university in Germany had their hand grip strength measured (in kilograms force (kgf)). The results are illustrated in Figure 8.1.

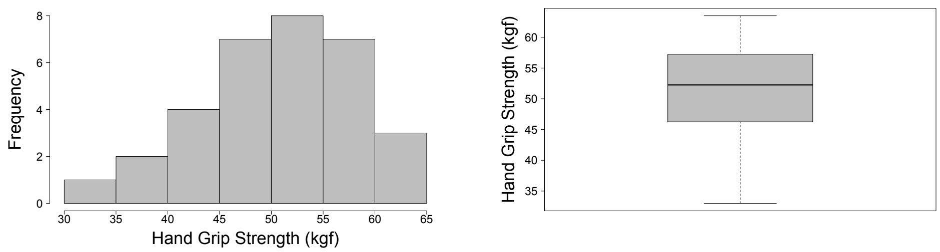


Figure 3.12: Plots of the grip strengths of 32 male student volunteers.

(a) Describe the distribution of the grip strengths.
(b) Estimate the 90th percentile of the grip strengths in this data set.
(c) Estimate the 90th percentile of the grip strengths of adult Germans.
(d) For this data, is the mean or the median the more appropriate measure of central tendency?
(e) Give rough estimates of the mean and median.
(f) Give a rough estimate of the standard deviation.

49. Tantius et al. (2014) investigated tensile properties of human umbilical cords. In one part of the study, 23 human umbilical cords were stretched until they broke, and their elongation (% increase in length) at the breaking point was recorded. (The breaking point of umbilical cords is of interest in forensic science, as mothers accused of killing a newborn might claim that an umbilical cord broke during childbirth and the newborn baby bled to death.) The elongation percentages are illustrated in Figure 3.13.

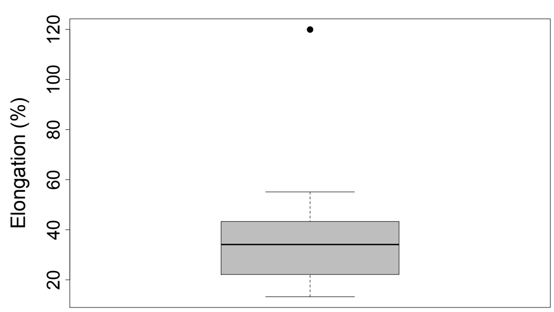


Figure 3.13: Boxplot of elongation percentage at breaking point.

(a) Describe the distribution of the elongation percentages.

(b) The mean elongation percentage of these 23 umbilical cords is 36.6. Estimate the mean if the outlier were to be removed from the calculations.
(c) Should the outlier be omitted from the analysis?
(d) The study actually involved 25 umbilical cords, but in two cases the machine did not recognize the break and thus did not record the elongation at the breaking point. These data points were omitted in this plot and in the calculation of the mean. How might this distort the results?

50. Harlioglu et al. (2012) investigated several characteristics of crayfish in a freshwater lake in Turkey. In this study, twenty-five adult male *Astacus leptodactylus* crayfish were sampled, and several variables were measured. One of the variables was body weight (in grams). The weights are illustrated in Figure 3.14.

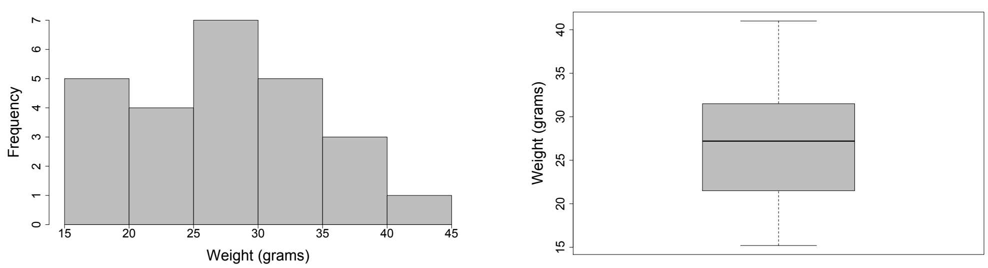


Figure 3.14: Weights of 25 adult male crayfish.

(a) Give rough estimates of the mean and median of the 25 weights.
(b) Suppose the researchers wished to estimate the average weight of all adult male crayfish in this lake. Would the mean of this sample be a reasonable estimate of this quantity?

51. The following observations represent lengths, in millimetres, of a sample of 16 *Heliconia bihai* flowers.

47.12	46.75	46.81	47.12	46.67	47.43	46.44	46.64
48.07	48.34	48.15	50.26	50.12	46.34	46.94	48.36

Summary statistics: $\bar{x} = 47.5975$, median $= 47.12$, and $s = 1.212878$.

Suppose we want to change the measurement scale, and have the measurements represent the number of inches the length is in excess of 40 mm. To carry this out, we (correctly) subtract 40, then divide the result by 2.54.

(a) What is the mean of the new values?
(b) What is the median of the new values?
(c) What is the standard deviation of the new values?
(d) What is the variance of the new values?

52. A person is weighing rocks in a lab in the United States. They are new to the United States from Canada, and they have been taking all of the measurements in kilograms. The sample mean and standard deviation of the weight of the rocks (in kilograms) were found to be 15.10 and 1.70, respectively.

    At this point the person realizes two important points: 1) The scale wasn't calibrated properly, and every measurement in the data set should have 2.00 kilograms added to it. 2) Their boss wants the mean and the standard deviation in pounds, not kilograms (1 kg = 2.20 pounds). The person correctly realizes that they must add 2.00 to all of their measurements, then multiply the result by 2.20. What are the mean and standard deviation of the correct weights in pounds?

Chapter 4

Probability

J.B.'s strongly suggested exercises: 2, 4, 5, 6, 7, 8, 14, 16, 17, 21, 26, 30, 35, 39, 43, 49, 57, 61

4.1 Introduction

4.2 Basics of Probability

4.2.1 Sample Spaces and Sample Points

1. Two cards are drawn without replacement from a well-shuffled deck.

   (a) What is the sample space of this experiment?
   (b) What are the the sample points? How many sample points are there?
   (c) Are the sample points equally likely?
   (d) What is the probability of getting a pair of twos?

2. Two balanced six-sided dice are rolled, and the numbers that come up are recorded.

   (a) What is the sample space of this experiment?
   (b) What are the the sample points? How many sample points are there?
   (c) Are the sample points equally likely?
   (d) What is the probability of rolling a sum of 2? (Rolling two dice such that the sum of the values on the up face is 2.)
   (e) What is the probability of rolling a sum of 7? (Rolling two dice such that the sum of the values on the up face is 7.)

4.3 Rules of Probability

4.3.1 The Intersection of Events

4.3.2 Mutually Exclusive Events

4.3.3 The Union of Events

4.3.4 Complementary Events

Questions for the intersection of events, mutually exclusive events, the union of events, and complementary events:

3. For each of the following events, draw a Venn diagram and shade the region that represents it.
   (a) $A \cup B^c$
   (b) $A \cap B^c$
   (c) $A^c \cup B^c$
   (d) $A^c \cap B^c$
   (e) $(A \cup B)^c$

4. Suppose $P(A) = 0.6$, $P(B) = 0.3$, and $P(A \cap B) = 0.2$.
   (a) What is $P(A^c)$?
   (b) What is $P(A \cup B)$?
   (c) What is $P((A \cup B)^c)$?
   (d) Are A and B mutually exclusive?
   (e) Are A^c and B^c mutually exclusive?

5. The following table gives the distribution of blood types in Canada.[1]

	A	B	AB	O
Rh positive	0.36	0.076	0.025	0.39
Rh negative	0.06	0.014	0.005	0.07

Table 4.1: Proportions of Canadians with different blood types. Every person falls into one and only one of these 8 categories.

Suppose we randomly select a Canadian. Consider the following events:

[1] As given on the Canadian Blood Services website. Accessed January 9, 2014.

A: The person has blood type A.
B: The person has blood type B.
C: The person has blood type AB.
D: The person has blood type O.
E: The person is Rh positive.

(a) Symbolically, what is the event that the person has type O negative blood? (If a person has type O blood and a negative Rh factor, it is said that they have type O negative blood.) What is the probability of this event?
(b) What is the probability the person has blood type O or a positive Rh factor?
(c) Which pairs of events are mutually exclusive?
(d) A person with A negative blood can receive a blood transfusion from someone who also has type A negative blood, or from someone who has type O negative blood. In symbols, what is the event that a randomly selected Canadian can donate to someone with type A negative blood? What is the probability of this event?

Conditional Probability

Independent Events

Questions for conditional probability and independent events:

6. Suppose $P(A) = 0.60$, $P(B) = 0.20$ and $P(A \cup B) = 0.65$.

   (a) What is $P(A \cap B)$?
   (b) What is $P(A|B)$?
   (c) What is $P(B|A)$?
   (d) Are A and B independent?

7. A balanced six-sided die is about to be rolled once.

   - Let A be the event that the number on the top face is greater than or equal to 4.
   - Let B be the event that the number on the top face is even.
   - Let C be the event that the number on the top face is a 6.

   (a) What is $P(B|C)$
   (b) What is $P(C|B)$
   (c) What is $P(C|B^c)$
   (d) What is $P((B \cap C)|A)$?
   (e) Which, if any, of the 3 pairs of events $(A, B), (A, C), (B, C)$ are independent?

The Multiplication Rule

8. Cards are drawn without replacement from a well shuffled deck.
   (a) If two cards are drawn, what is the probability both cards are hearts?
   (b) If two cards are drawn, what is the probability both cards are fives?
   (c) If four cards are drawn, what is the probability all four cards are hearts?

9. Two balanced six-sided dice are rolled. (The rolling is done in such a fashion that different rolls can be considered independent.)
   (a) What is the probability both dice have a four on the up face?
   (b) What is the probability that an even number is rolled on the first die, and a number greater than four is rolled on the second?

4.4 Examples

10. If $P(A) = 0.80$, $P(B) = 0.40$, and $P(B|A) = 0.25$, what is $P(A|B)$?

11. Suppose we know the following about events A, B, and C:

$$P(A) = 0.42,\ P(B) = 0.44,\ P(C) = 0.40$$

$$P(A \cap B) = 0.16,\ P(A \cap C) = 0.20,\ P(B \cap C) = 0.12$$

$$P(A \cap B \cap C) = 0.04$$

   (a) Draw a Venn diagram and fill in the appropriate probabilities.
   (b) What is $P(A \cup B)$?
   (c) What is $P(A \cup B \cup C)$?
   (d) What is $P(A|B \cap C)$?

12. The following table is based on a 2009 poll in the U.S., in which 1000 randomly selected adults were asked if they approved or disapproved of the way Barack Obama is handling his job as president.

	Republicans	Democrats	Independents
Approve	60	361	139
Disapprove	270	49	121

   (a) If a person is randomly selected from this group of 1000, what is the probability they are Republican?

(b) What is the probability they said they approve of the way Obama is handling his job?

(c) Given a randomly selected person is Republican, what is the probability they approve?

13. Consider the Venn diagrams shown in Figure 4.1. For each plot, assume that the area of the square is 1, and that the area of each region is equal to its probability of occurring. The smaller circle represents event A, and the larger circle represents event B.

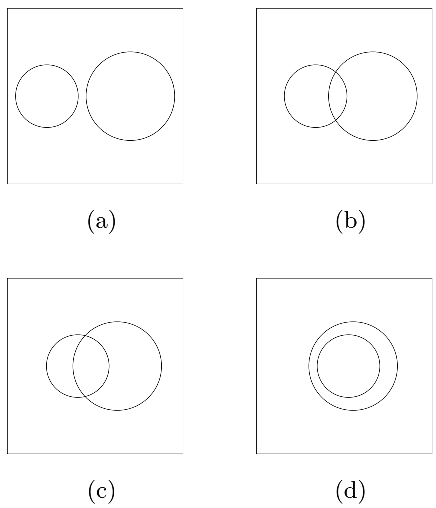


Figure 4.1: Venn diagrams for Question 13.

(a) One of the plots was created for the situation where A and B are independent. Which one is it?

(b) For each plot, state whether $P(A|B) = P(A)$, $P(A|B) > P(A)$, or $P(A|B) < P(A)$.

4.5 Bayes' Theorem

14. An airport security official tests an experimental electronic sniffing device that can detect explosive materials. All passengers are scanned by this device, which sounds an alarm if it detects explosive material. The device is not perfectly reliable. It will falsely sound an alarm 2% of the time if the person being scanned is not carrying explosive material It will also fail to sound an alarm 8% of the time if the person being scanned is carrying explosive material.

Assume that 1 in 10,000 passengers carries explosive material. If the device scans a randomly selected passenger:

(a) What is the probability the alarm sounds?
(b) If the alarm sounds, what is the probability the person is actually carrying explosive material?

15. A small regional airport is serviced by 2 commercial airlines. Caribou Air accounts for 65% of commercial flight arrivals. Arrivals from this airline are late 22% of the time. Polar Express Air accounts for 35% of commercial flight arrivals. Arrivals from Polar Express Air are late 42% of the time.

    (a) If a commercial flight arrival is randomly selected, what is the probability it is late?
    (b) Given a randomly selected commercial flight arrival is late, what is the probability the flight is a Caribou Air flight?

4.6 Counting rules: Permutations and Combinations

16. (a) At the start of a Texas holdem poker hand, each player is dealt two cards from a standard 52 card deck. How many possible two-card hands are there?
    (b) At the start of a five card draw poker hand, each player is dealt five cards from a standard 52 card deck. How many possible five-card hands are there?
    (c) What is the probability of being dealt a flush (5 cards all of the same suit) in five card draw? (In this question, any hand in which all 5 cards are the same suit is considered to be a flush. In poker, straight flushes are considered to be distinctly different from a flush, but they are a flush for the purposes of this question.)

17. Suppose we are about to draw a simple random sample of 4 students from a class of 200.

    (a) How many different samples are possible?
    (b) What is the probability that any one of the individual 200 students gets selected in the sample?

18. (a) A committee of 5 people is about to be drawn from a group of 25 faculty. How many different groups of 5 people can be chosen for the committee?
    (b) A bussing company needs to assign bus drivers to 7 different routes. There are 30 bus drivers to choose from. How many different ways are there of assigning 7 bus drivers to the different routes?

4.7 Probability and the Long Run

19. Tom and Pete are wasting time by repeatedly betting \$1 on each toss of a fair coin. If the coin comes up heads, Tom wins \$1. If the coin comes up tails, Pete wins \$1. Which of the following statements are true?

    (a) The proportion of times heads has come up will tend toward 0.5 as the number of tosses tends to infinity.
    (b) The amount of money the losing player owes the winning player will tend to 0 as the number of tosses tends to infinity.
    (c) The probability they are exactly even after 1,000,000 tosses is greater than the probability they are exactly even after 10 tosses.

4.8 Chapter Exercises

4.8.1 Basic Calculations

20. Suppose $P(A) = 0.30$, $P(B) = 0.80$, and $P(A \cap B) = 0.125$.

    (a) What is $P(A \cup B)$?
    (b) What is $P((A \cup B)^c)$?
    (c) What is $P(A|B)$?
    (d) What is $P(B|A)$?
    (e) Are A and B independent?
    (f) What is $P(A^c \cup B)$

21. Suppose A and B are two events such that $P(A) = 0.5$, $P(B) = 0.2$ and $P(A \cup B) = 0.68$.

    (a) What is $P(A \cap B)$?
    (b) What is $P((A \cup B)^c)$?
    (c) What is $P(A|B)$?
    (d) What is $P(B|A)$?
    (e) Are A and B independent?
    (f) What is $P(A^c \cup B^c)$
    (g) What is $P(A^c \cap B^c)$

22. Suppose $P(A) = 0.2$, $P(B) = 0.2$, $P(A \cap B) = 0.2$. Which of the following statements are true?

    (a) A and B are mutually exclusive.

(b) A and B are independent.
(c) $P(A|B) = 0$.
(d) $P(A|B) = 1$.
(e) A and B^c are not independent.
(f) $P(A \cup B) = P(A \cap B)$.
(g) $P(A|B) = P(B|A)$.

23. Suppose a random sample of 100 students who have taken both course A and course B yielded the following:

	Passed Course B	Failed Course B
Passed Course A	71	19
Failed Course A	2	8

(a) If one of these students is randomly selected, what is the probability they passed course A?
(b) If one of these students is randomly selected, what is the probability they passed course A, if we know the student passed course B?
(c) If one of these students is randomly selected, are the events *the student passed course A* and *the student passed course B* independent?

24. Suppose $P(A) = 0.4$, $P(B) = 0.7$, and the probability that A and B both occur is 0.2.

(a) What is $P(A|(B \cup A)^c)$?
(b) What is $P(A \cap B|A \cup B)$?

25. A balanced six-sided die is rolled once.

(a) What is the probability of rolling an even number on the top face?
(b) What is the probability of rolling a 3?
(c) What is the probability the number on the top face is a 3, given it is an even number?
(d) What is the probability the number on the top face is at least 4, given it is an even number?
(e) Are the events *the number on the face is even* and *the number on the face is more than 3* independent?
(f) What is the conditional probability the number on the face is 2 or less, given the number is at least 4?

4.8.2 Concepts

26. Probabilities are often expressed in terms of **odds**. For example, we might say that the **odds in favour** of an event are 2:7 or that the **odds against** an event are 6:1. The *odds in favour* of an event is a ratio of two whole numbers that represents the ratio of the probability of the event to the probability of its complement. The *odds against* an event switches the order of the numbers (odds against represents the ratio of the probability of the event's complement to the probability of the event). For example, suppose we are about to roll a fair die once and we are hoping to roll a 2 (which has a probability of $\frac{1}{6}$). The odds in favour of rolling a 2 are 1:5. (On average, for every 2 that we roll, we will roll a different number 5 times.) The odds against rolling a 2 are 5:1.

    (a) Suppose we are about to roll a single die once. What are the odds in favour of rolling a 5 or a 6? What are the odds against rolling a 5 or a 6?
    (b) Suppose we are about to draw a single card from a well-shuffled deck. What are the odds in favour of drawing a jack? What are the odds against drawing a jack?
    (c) If the odds in favour of an event are a:b, then on average, for every a occurrences of the event, the event will not happen b times, and thus the probability of the event is $\frac{a}{a+b}$. If the odds in favour of an event are 5:3, what is the probability of this event?
    (d) If the odds against an event are 19:1, what is the probability of the event?

27. The following two identities are called De Morgan's laws, and they frequently arise in probability and logic. Verify these identities using Venn diagrams.

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

28. We have already been introduced to the addition rule for two events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Use an argument based on Venn diagrams to come up with the addition rule for three events. That is, express $P(A \cup B \cup C)$ in terms of the probabilities of A, B, and C and the probabilities of their intersections.

29. Which one of the following statements is true? (Assume for the purposes of this question that the events have non-zero probabilities of occurring.)

    (a) Mutually exclusive events can be independent.
    (b) Mutually exclusive events are always independent.
    (c) Independent events are never mutually exclusive.
    (d) Independent events are always mutually exclusive.

30. Which of the following statements are true? (There may be more than one true

statement; check all that are true.)

(a) If $P(A) = 0.30$ and $P(B) = 0.30$, then $P(A \cap B) = 0.09$.
(b) If $P(A) = 0.30$ and $P(B) = 0.30$, and $P(A \cup B) = 0.60$, then A and B are independent.
(c) If $P(A) = 0.50$, and the probability that B occurs given A occurs is 0.50, then $P(A \cap B) = 0.25$.
(d) If $P(A) = 0$, then A is independent of any event.
(e) If $P(B|A) = 0.80$, then $P(B) > 0.80$.

31. Which of the following statements are true? (There may be more than one true statement; check all that are true.)

    (a) If $P(A) = 0.6$, $P(B) = 0.7$, then $0.3 \leq P(A \cap B) \leq 0.6$.
    (b) If $P(A) = 0.5$, then for any event B: $0 \leq P(A \cap B) \leq 0.5$.
    (c) If $P(A|B) = 0$, then A and B are mutually exclusive.
    (d) Suppose that $P(A) = 0.99$. Event A and its complement are mutually exclusive and independent.
    (e) If $P(A|B) \neq P(B)$ then A and B are not independent.

32. Two events A and B are mutually exclusive. If $P(A) = 0.5$ and $P(B) = 0.3$, are A and B independent?

33. Suppose a fair coin is tossed twice, in such a fashion that the two tosses are independent. Let A be the event that heads comes up on the first toss, and B be the event that heads comes up on the second toss. Which of the following 3 pairs of events represent independent events.

    (a) A and B
    (b) A and B^c
    (c) A and A^c

34. Consider the diagrams shown in Figure 4.2. For each plot, assume that the area of the square is 1, and that the area of each region is equal to its probability of occurring. For which plots, if any, are A and B independent?

35. Suppose we randomly select a Canadian adult. Let A be the event that the person is male. Let B be the event that the person weighs more than 100 kg. Which of the following statements are true?

    (a) $P(A|B) < P(A)$.
    (b) $P(B|A) < P(B)$.
    (c) A and B are mutually exclusive.
    (d) A and B are not independent.

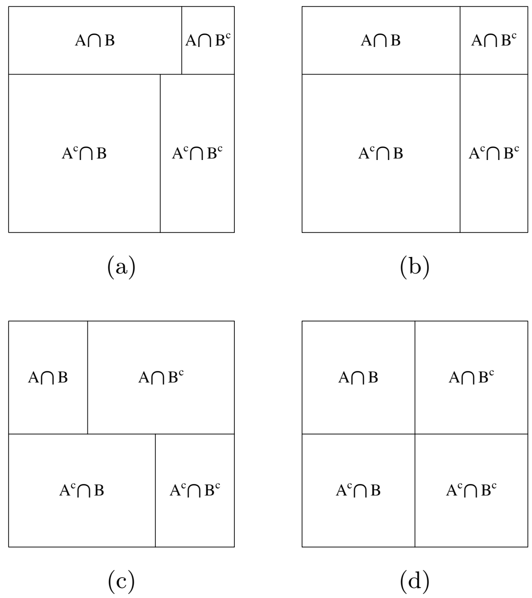


Figure 4.2: Venn diagrams for Question 34.

36. For the following pairs of events, state whether $P(B|A) = P(B)$, $P(B|A) > P(B)$, or $P(B|A) < P(B)$.

    (a) For your current location:
    A: It rains on Tuesday of next week.
    B: It rains on Wednesday of next week.

    (b) For your current location:
    A: It rains on Tuesday of next week.
    B: It rains on the day that is 10 years from today.

    (c) For a randomly picked Canadian adult:
    A: The person did not finish high school.
    B: The person has been to prison.

37. Two inspectors, A and B, work side by side at a factory. Their jobs involve inspecting the quality of each part that is produced and giving each part a pass or fail rating. They both just started the job, and are both very lazy. They decide to randomly choose between passing and failing each part. For each part, one of them tosses a coin twice and they both make their decisions based on these 2 tosses. Inspector A passes the part if the first coin comes up heads and fails it otherwise. Inspector B passes the part if heads come up exactly once and fails it otherwise. Consider the events:

    A: Inspector A passes the part. B: Inspector B passes the part.

(a) Are A and B mutually exclusive?
(b) Are A and B independent?

38. Daniel has to do an online assignment for one of his classes. The assignment consists of 20 multiple choice questions, where there are 5 answer options for each question. Daniel knows nothing about the material, and if he were to do the assignment on his own he would guess randomly on each question. Daniel's friend Magnus has an 80% chance of getting a passing grade on the assignment. (Magnus works a *little* harder than Daniel.) Daniel is tempted to cheat on the assignment by copying his friend Magnus's answers, but Daniel doesn't like cheating, and he knows that the university considers cheating on an assessment to be a serious case of academic misconduct. Daniel is torn between cheating and doing the assignment on his own, so he elects to let the fates decide. He will toss a coin once, and if it comes up heads he will copy every answer off his friend Magnus. If the coin comes up tails, Daniel will do the assignment himself, guessing at every answer.

(a) Estimate the probability that Daniel passes the assignment. (We haven't learned how to calculate the exact probability of this yet, but using a simple argument we should be able to come pretty close.)
(b) Are the events *Daniel gets a passing grade on the assignment* and *Magnus gets a passing grade on the assignment* mutually exclusive?
(c) Are the events *Daniel gets a passing grade on the assignment* and *Magnus gets a passing grade on the assignment* independent?

39. 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) Probabilities can be negative.
(b) For any two events A and B, $P(A|B) > P(A)$.
(c) If A and B are independent, then $P(A|B) = 0$.
(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*.
(e) If two events A and B are independent, then their complements (A^c and B^c) are also independent.
(f) If A and B are mutually exclusive, then $P(A|B) = 0$.

40. 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) If $P(A) = P(B)$, then A and B are independent.
(b) If $P(A) > P(B)$, then $P(A|B) \geq P(B|A)$.
(c) If A and B are independent, then $P(A \cup B) = P(A) + P(B)$.

(d) If A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.
(e) If $P(A|B) = 1$, then $P(B|A) = 1$.

4.8.3 Applications

41. In a 2008 study that was picked up by major news networks, two Manhattan teenagers investigated whether restaurants and grocery stores that served sushi were mislabelling the fish that was used. They sent samples to the University of Guelph for analysis, and found that 25% of the fish samples that could be genetically identified were mislabelled. Often, cheaper fish was served instead of a more expensive variety. Assume for the purposes of this question that a randomly selected sushi dish (that contains fish) at a randomly selected restaurant has a 25% chance of having mislabelled fish.

    (a) Suppose you go to a randomly selected sushi restaurant. You order 4 randomly selected types of sushi (the fish variety). Assuming independence, what is the probability that you get at least one piece that is mislabelled?
    (b) True or False? The independence assumption in the previous question is perfectly justified.

42. Suppose a group of 50 students is made up of 20 males and 30 females. Twenty five of these students have part time jobs. 15 of the males do not have a job. Suppose one of these 50 students is randomly selected.

    (a) What is the probability they do not have a job?
    (b) What is the probability they are a male that does not have a job?
    (c) If they have a part-time job, what is the probability they are male?
    (d) If they are a female with a job, what is the probability they are male?
    (e) For these 50 students, is job status independent of gender?

43. Is there a relationship between education and smoking status among French men? A study[2] measured several variables on 459 healthy men in France who were attending a clinic for a check up. Table 4.2 gives the results of the study. Suppose one of these 459 men is randomly selected.

    (a) What is the probability the man is a smoker?
    (b) What is the probability the man has a university education?
    (c) Given the man is a smoker (moderate or heavy), what is the probability that he has a university education?

[2]Marangan et al. (1998). Diet, antioxidant status, and smoking habits in French men. *American Journal of Clinical Nutrition*, 67:231–239.

	Nonsmoker	Ex-smoker	Moderate	Heavy
Primary School	56	54	41	36
Secondary School	37	43	27	32
University	53	28	36	16

Table 4.2: Smoking status and level of education for 459 French men.

(d) Given the man is a smoker (moderate or heavy), what is the probability that he has at least a high school education?

(e) Given the man is a nonsmoker, what is the probability that he has less than a high school education?

(f) Given the man has only a secondary school education, what is the probability that he is a smoker?

(g) Given the man has at least a secondary school education, what is the probability that he is a smoker?

(h) Are they events *the man has a university education* and *the man is a smoker* independent?

(i) Are they events *the man has a university education* and *the man is a smoker* mutually exclusive?

44. Is there a relationship between fatty fish consumption and the rate of prostate cancer? A study[3] followed 6272 Swedish men for 30 years. They were categorized according to their fish consumption, and to whether they developed prostate cancer. The following table summarizes the results. Suppose one of these 6272 men is randomly selected.

	Fish consumption			
	Never/Seldom	Small	Moderate	Large
Prostate cancer	14	201	209	42
No prostate cancer	110	2420	2769	507

(a) What is the probability their fish consumption was large?

(b) What is the probability their fish consumption was small, moderate, or large?

(c) What is the probability they got prostate cancer?

(d) What is the probability they got prostate cancer, given their fish consumption was never/seldom?

(e) What is the probability they got prostate cancer, given their fish consumption was large?

(f) What is the probability their fish consumption was large, given they got prostate cancer?

(g) What is the probability they got prostate cancer, given their fish consumption

[3] Terry et al. (2001). Fatty fish consumption and risk of prostate cancer. *Lancet*, 357:1764–1766.

was small or moderate?

45. The breakdown of male and female full-time undergraduate students at the University of Guelph, University of Waterloo, and Wilfrid Laurier University in the fall semester of 2009 is given in Table 4.3.

University	Male	Female	Total
Guelph	6,366	10,417	16,783
Waterloo	13,770	10,607	24,377
Laurier	4,803	6,254	11,057
Total	24,939	27,278	52,217

Table 4.3: Undergraduate enrolment in the fall semester of 2009.

If one of these 52,217 students is randomly selected:

(a) What is the probability they attended the University of Guelph that semester?
(b) What is the probability they are female?
(c) What is the probability they attended the University of Guelph, given they are female?
(d) What is the probability they are female, given they attended the University of Guelph?
(e) What is $P(F|G \cup L^c)$, where F represents the event that the randomly selected student is female, G represents the event that the randomly selected student attended the University of Guelph that semester, and L represents the event that the randomly selected student attended Wilfrid Laurier University that semester?

46. A restaurant conducts a survey of customers as they are leaving. The following table gives a summary of the results.

	Dining experience was good	Dining experience was not good
Men	23	7
Women	17	18

(a) If one of these 65 customers is selected at random, what is the conditional probability the dining experience was good, given they are male.
(b) If one of these 65 customers is selected at random, what is the conditional probability they are a man, given they are a woman.
(c) If one of these 65 customers is randomly selected, are the events *the customer is a man* and *the dining experience was good* independent?
(d) Are the events *the customer is a man* and *the customer is a woman* independent? Are they mutually exclusive?

47. A student is worried about sleeping in and missing their final exam, so they set two battery powered alarm clocks. The probability the first alarm will go off is 0.99, and the probability the second alarm will go off is 0.98. Assume that these clocks work independently. (Independence might be a sketchy assumption in this spot, but let's go with it.)

    (a) What is the probability that both alarms go off?
    (b) What is the probability that neither alarm will go off?
    (c) What is the probability that at least one alarm goes off?
    (d) Which one of the above is the most important question, given the nature of the problem?

48. Two cards are drawn *with replacement* from a standard deck.

    (a) What is the probability that both cards are kings?
    (b) Given there is at least one king drawn, what is the probability both cards are kings?
    (c) What is the probability that exactly one card is red?
    (d) What is the probability that at least one card is red?
    (e) What is the probability at least one card is a king, and at least one card is red?

49. Two cards are drawn *without replacement* from a standard deck.

    (a) What is the probability that both cards are kings?
    (b) Given there is at least one king drawn, what is the probability both cards are kings?
    (c) What is the probability that exactly one card is red?
    (d) What is the probability that at least one card is red?
    (e) What is the probability at least one card is a king, and at least one card is red?

50. Mateo is both rich and a horrible driver. He is going to drive to a convenience store to buy cigarettes, and he must decide which one of his two cars he will take. He decides to flip a fair coin twice, and if it comes up heads both times he will take his Ferrari, and otherwise he will take his Lexus. If he takes his Ferrari he has a 0.08 probability of getting into an accident. If he takes his Lexus, the probability that he gets into an accident is 0.04.

    (a) What is the probability that Mateo gets into an accident?
    (b) Given he takes his Ferrari, what is the probability that he gets into an accident?
    (c) Given he gets into an accident, what is the probability he took the Ferrari?

51. At a certain gas station, 80% of the people buy regular gas, 5% buy midgrade, and

15% buy premium. Suppose if a randomly selected person pumps regular gas, there is a probability of 0.02 that they leave without paying. If they pump midgrade, the probability is 0.04, and if they pump premium the probability is 0.14.

(a) For a randomly selected gas pumper, what is the probability they leave without paying?
(b) If they leave without paying, what is the probability that they pumped premium gas?

52. You need to buy a part from a supplier. The supplier has 3 different sources for this part. The supplier will get the part from Source A with probability 0.55, from Source B with probability 0.35, and from Source C with probability 0.10. If the part comes from Source A, there is a 40% chance it will last for a year. If the part comes from Source B, there is an 90% chance it will last for a year. If the part comes from Source C, there is a 55% chance it will last for a year. You order a part from the supplier.

(a) What is the probability the part lasts for a year?
(b) Given the part lasts for a year, what is the probability that it came from Source A?

53. In sample surveys with sensitive questions, respondents are likely to lie, and thus the survey results would not reflect reality. (For example, it may be difficult to get a truthful answer to the question, "have you ever fantasized your spouse was dead?") One way to combat this effect is to implement a **randomized response** survey. For example, in a telephone survey an individual may be told to toss a coin before responding. If the coin comes up heads, they should respond "Yes", regardless of the truth. If coin comes up tails they should tell the truth. Incorporating a randomized response adds a level of variability to the responses, which is undesirable. But respondents may be more willing to answer truthfully, since the person on the other end of the telephone will not know if the respondent is telling the truth, or if they responded that way simply because the coin fell a certain way.

(a) Suppose that in reality, 20% of a population has stolen from their employer. What is the probability a randomly selected person from this population responds "Yes" to the question "Have you stolen from your employer?" (Assuming the person follows the directions given above.)
(b) Assume again that 20% of a population has stolen from their employer. If a randomly selected person responds "Yes", what is the probability they have stolen from their employer?
(c) Suppose the probability that a person has stolen from their employer is not known (which is typically the case). If 67% of the respondents answer "Yes" to the question, what is the best estimate of the proportion of individuals who have stolen from their employer?

54. Participants in a survey were asked to use a *randomized response* to the question "have you ever cheated on your spouse?" They were asked to roll a die, and if the die came up with a 5 or a 6 they were to tell the truth. If the die came up with a 1, 2, 3, or 4 they were to answer "yes", regardless of the truth. If 82% of the respondents answer "yes" to the question, what is the best estimate of the proportion of these people who have actually cheated on their spouse?

55. In a certain population of mice, each mouse has a one-in-a-thousand chance of having a specific genetic disorder. You intend to sample mice and check to see if they have this disorder. Suppose that the test for this disorder works perfectly, and the mice can be considered independent.

    (a) If you sample 30 mice, what is the probability that none of them have the genetic disorder?
    (b) What is the probability that more than 500 mice will be sampled before finding the first mouse with the disorder?
    (c) How many mice would need to be sampled before there is at least a 50% chance of finding at least one with the disorder?

56. You have an expensive system that relies on an important component. This component has a probability of failure of 0.32. You feel this probability is much too high, and you have designed a system in which you can install as many independent components as you like. The system connects these devices in parallel, so if at least one of them works the system works.

    (a) If you connect 4 of these components in parallel, what is the probability of failure of the system?
    (b) What is the smallest number of these components that you would have to connect in parallel for the probability of failure to be less than 0.0001?

57. An important satellite guidance system relies on a very fragile component that has a 0.77 probability of failure. Since this failure probability is very high, designers want to put in independent back-up components (each of these also has a probability of failure of 0.77). They wish to connect them in parallel such that the guidance system works as long as at least one of the components works.

    (a) If you connect 5 of these components in parallel, what is the probability the system fails?
    (b) If you connect 10 of these components in parallel, what is the probability the system fails?
    (c) What is the smallest number of components that must be connected in order for the probability the system works to be at least 0.99990?

58. You intend to do research into people with a certain rare genetic marker. The

probability a randomly selected person has the marker is 0.0001. You need people with the marker for your experiment. Assume that you can randomly sample people from the population.

(a) What is the probability the first person with the marker occurs on the 2nd person sampled?
(b) What is the probability the first person with the marker occurs on the 10th person sampled?
(c) What is the probability that the first person with the marker occurs on or before the 2,000$^{\text{th}}$ person sampled?

59. A large art museum has two painting experts on staff. One of their jobs is to assess whether given artwork is authentic (they decide whether it is genuine or a forgery). Expert A and Expert B both look at each artwork individually, and arrive at their own conclusions. A summary of the results of inspections of 200 pieces of art follows.

	Judged genuine by B	Judged a forgery by B
Judged genuine by A	174	4
Judged a forgery by A	2	20

(a) If one of these pieces of art is randomly selected, what is the probability that Expert A thought it to be genuine?
(b) One of these pieces of art is randomly selected. Given Expert B thought it a forgery, what is the probability that Expert A also thought it a forgery?
(c) Are events *Expert A thinks the piece is a forgery* and *Expert B thinks the piece is a forgery* mutually exclusive? Are they independent?

60. John has a crush on Stephanie. A friend of John's is having a party on Friday night. If Stephanie goes to the party, the probability that John will also go is 0.94. If Stephanie does not go to the party, the probability that John goes is 0.03.

(a) If the probability that Stephanie does not go to the party is 0.20, what is the probability that John goes?
(b) Suppose now that the probability Stephanie goes to the party is unknown, but the conditional probabilities that John goes (0.94, 0.03) are still the same. If the probability that John goes to the party is 0.38, what is the probability that Stephanie goes to the party?
(c) Are the events *John goes to the party* and *Stephanie goes to the party* mutually exclusive? Are they independent?

61. A certain genetic defect affects 0.001% of the population. A test is available for this defect. If someone has the defect, the test will fail to detect it with probability 0.01. If somebody does not have the defect, the test will give a false positive with

probability 0.005.

(a) What is the probability that a randomly selected person tests positive?
(b) Given a randomly selected person tests positive for this defect, what is the probability they have the defect?

62. The following table is based on a Statistics Canada publication.[4]

40 years	0.99864
41 years	0.99852
42 years	0.99837
43 years	0.99818
44 years	0.99798
45 years	0.99775
46 years	0.99751
47 years	0.99726
48 years	0.99702
49 years	0.99678

The table represents part of a **life table**, which gives various estimated probabilities of survival. The probabilities given here are estimated conditional probabilities. For example, the estimated probability that an exactly 40 year old Ontario man survives until his 41^{st} birthday is .99864. The estimated probability that an exactly 41 year old Ontario man survives until his 42^{nd} birthday is .99852, etc.

(a) Based on this table, what is the probability an Ontario man, alive on his 47th birthday, survives until his 48th?
(b) Based on this table, what is the probability an Ontario man, alive on his 47th birthday, dies before his 48th?
(c) Based on this table, what is the probability a male child born in Ontario lives until their 40th birthday?
(d) Based on this table, what is the estimated probability that an Ontario man who has survived until his 45^{th} birthday dies when he is 48? That is, he reaches his 48^{th} birthday but dies before his 49^{th}.
(e) What is the estimated probability that a randomly selected Ontario man, alive on his 45th birthday, survives until his 48th?

[4] The full table is available at `http://www.statcan.gc.ca/pub/84-537-x/t/txttables/ontm.txt`. Accessed January 11 2014.

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. For each of the following random variables, state whether they are discrete or continuous.
   (a) The number of poker hands dealt in a casino in an hour.
   (b) The time until completion of a randomly selected poker hand in a casino.
   (c) The amount of cheese on a randomly selected cheeseburger at a fast food restaurant.
   (d) The number of moves in a game at the World Chess Championship.
   (e) The duration of a match at the World Chess Championship.
   (f) The number of die rolls required to roll a six 18 times in a row.
   (g) The sum of the numbers that come up on the top face when a die is rolled 400 times, divided by 3617.

5.3 Discrete Probability Distributions

2. Suppose that two cards are drawn without replacement from a standard 52 card deck. Let X represent the number of hearts drawn. Find the probability distribution of X.

3. Consider the following probability distribution of a random variable X.

x	10	20	30	40
$p(x)$	0.2	0.2	0.5	?

   (a) Is X discrete or continuous?
   (b) What is the value of the missing probability?
   (c) What is the most likely value of X?
   (d) What is $P(X < 32)$?
   (e) What is the conditional probability X is less than 25, given X is less than 35?

4. Which of the following are valid discrete probability distributions?

(a)

x	-10	1.7	30000
$p(x)$	0.2	0.2	0.6

(b)

x	1	2	3
$p(x)$	0.2	0.2	0.2

(c)

x	0.1	0.2	0.3	4000	500000
$p(x)$	0.2	0.2	0.2	0.2	0.2

(d)

x	1	2	3	4	5
$p(x)$	-0.2	0.3	0.3	0.4	0.2

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. Consider the following probability distribution of a random variable X.

x	-2.5	0	2.5	5.0	7.5
$p(x)$	0.10	0.10	0.10	0.40	?

(a) Is X discrete or continuous?
(b) What is the value of the missing probability?
(c) What is the most likely value of X?
(d) What is $P(-1 < X < 4)$?
(e) What is the expected value of X?
(f) What is the standard deviation of X?

6. **Hellin's law** is a rough guideline that gives approximate probabilities of multiple births in pregnancies that are not the result of fertility treatments. Hellin's law states that approximately 1 in 89 pregnancies result in twins, approximately 1 in 89^2 pregnancies result in triplets, approximately 1 in 89^3 pregnancies result in quadruplets, and so on.

   The following table gives the approximate distribution of the number of babies delivered at birth for pregnancies that are not the result of fertility treatments. (The values in the table are roughly based on Hellin's law. Quintuplets and higher order multiple births are extremely rare and are ignored for the purposes of this question.)

Number of babies	1	2	3	4
Probability	0.988	0.01187	0.000128	0.000002

   For the following questions, "pregnancy" refers to a pregnancy that is not the result of fertility treatments.

   (a) What is the mean number of children per pregnancy?
   (b) What is the standard deviation of the number of children per pregnancy?
   (c) What is the probability of a multiple birth?
   (d) Given the pregnancy results in a multiple birth, what is the probability it results in twins?
   (e) Given the pregnancy results in triplets or quadruplets, what is the probability it results in triplets?

7. Which of the following statements are true?

   (a) Probabilities of impossible events can be negative.
   (b) If a random variable can take on an infinite number of possible values, then it cannot be a discrete random variable.
   (c) The mean of a discrete random variable cannot be negative.
   (d) The standard deviation of a discrete random variable cannot be negative.
   (e) The mean of a discrete random variable cannot be greater than its standard deviation.

5.3.1.2 Properties of Expectation and Variance

8. Suppose:

   - X is a random variable with $\mu_X = 8.7$ and $\sigma_X = 18.5$.
   - Y is a random variable with $\mu_Y = 14.9$ and $\sigma_Y = 21.0$.
   - X and Y are independent.

   (a) What is the mean of $X + Y$?
   (b) What is the standard deviation of $X + Y$?
   (c) What is the standard deviation of $X - Y$?

9. Twenty university students are taking both Course A and Course B. In course A the professor is very conscientious, and she makes an effort to make the tests consistent with and representative of the material covered in the class. Course B is taught by a professor who doesn't give any tests, and at the end of the semester he simply randomly assigns grades to students. Is a student's grade in course A independent of their grade in course B? Explain.

10. An assembly procedure consists of three independent steps. Step 1 has a mean time to completion of 6 minutes and a standard deviation of 1.2 minutes. Step 2 has a mean time to completion of 125 minutes and a standard deviation of 15.0 minutes. Step 3 has a mean time to completion of 17 minutes and a standard deviation of 3 minutes.

    (a) What is the mean time to completion of the entire procedure?
    (b) What is the standard deviation of the time to completion of the entire procedure?
    (c) Suppose we wish to measure time in hours instead of minutes. What are the mean and variance of the time (in hours) to completion?

5.4 The Bernoulli Distribution

11. Suppose a single card is drawn from a well-shuffled 52 standard deck.

    (a) What is the distribution of the number of eights that are drawn?
    (b) What is the mean number of eights that are drawn?
    (c) What is the standard deviation of the number of eights that are drawn?

5.5 The Binomial Distribution

12. Suppose X is a binomial random variable with parameters $n = 15$ and $p = 0.2$.

    (a) What is $P(X = 3)$?
    (b) What is $P(X \leq 3)$?
    (c) What is $P(X > 3)$?
    (d) What is $P(2 \leq X < 5)$?
    (e) What is $P(X = 3|X \leq 4)$
    (f) What is the mean of X?
    (g) What is the standard deviation of X?

13. According to the United States Centers for Disease Control and Prevention, approximately 3% of babies in the United States are born with major structural or genetic birth defects. Suppose 50 newborns in the United States are randomly selected.

    (a) What is the probability that exactly 2 have major structural or genetic birth defects?
    (b) What is the probability that no more than 2 have major structural or genetic birth defects?
    (c) What is the probability that exactly 10 have major structural or genetic birth defects?
    (d) What is the expectation of the number with major structural or genetic birth defects?
    (e) What is the standard deviation of the number with major structural or genetic birth defects?

14. Which of the following statements are true?

    (a) The variance of a binomial distribution is always greater than the mean.
    (b) A binomial random variable can take on negative values.
    (c) The mean of a binomial random variable can be negative.
    (d) The variance of a binomial random variable can be negative.
    (e) A binomial random variable is a discrete random variable.
    (f) Every discrete random variable is a binomial random variable.
    (g) A binomial random variable represents a count.
    (h) The mean of a binomial random variable is np.
    (i) The standard deviation of a binomial random variable is $\sqrt{np(1-p)}$.
    (j) For any given n, the variance of a binomial random variable is greatest when $p = 0.5$.

5.5.1 Binomial or Not?

15. Which of the following random variables have a binomial distribution?

    (a) The number of hearts if 10 cards are drawn without replacement from a standard deck.
    (b) The number of times heads comes up if a biased coin is tossed 10 times.
    (c) The number of putts a golfer holes in their next 10 attempts.
    (d) The number of newborn babies in a maternity ward on a randomly selected day.
    (e) The amount of money withdrawn by a randomly selected customer at a bank.

16. Which of the following random variables have a binomial distribution?

    (a) The number of hearts if 10 cards are drawn with replacement from a standard deck.
    (b) The number of fingers on a randomly selected newborn.
    (c) The number of times a sum of 7 is rolled when a pair of dice is rolled 14 times.
    (d) The number of children a randomly selected woman has.
    (e) The number of times a randomly selected person uses the washroom on a randomly selected day.
    (f) The weight of 15 randomly selected playing cards.

5.5.2 A Binomial Example with Probability Calculations

17. A friend of yours says they have extra-sensory perception (ESP), and claims they can pull the king of spades from a well shuffled 52-card deck approximately 20% of the time. You think they are being foolish and you decide to test their claim. You open a brand new deck of cards, shuffle it well, and ask them to draw the king of spades. You repeat this procedure until they have drawn a card 50 times. They manage to draw the king of spades twice in the 50 draws.

    (a) If your friend does not have ESP, and is merely randomly pulling a card each time, what is the probability they draw the king of spades at least twice in 50 draws?
    (b) Does this experiment give strong evidence that your friend does not have ESP?

5.6 The Hypergeometric Distribution

18. An urn contains 22 white and 18 red balls. Six balls are randomly chosen without replacement.

    (a) What is the probability that exactly 4 white balls are chosen?
    (b) What is the probability that no more than 1 white ball is chosen?
    (c) What is the mean number of white balls chosen? What is the mean number of red balls chosen?

19. You have an idea to start a business that delivers pizzas using stretch limousines. You need to hire 5 drivers, and 20 people apply for the positions. Unknown to you, 7 of these 20 people have criminal records. Suppose you (foolishly) decide to hire a random sample of 5 drivers from the 20 applicants.

    (a) What is the probability that exactly three of those hired have a criminal record?
    (b) What is the probability that exactly two of those hired have a criminal record?
    (c) What is the probability that at least one of those hired has a criminal record?
    (d) On average, how many of the 5 people you hire will have a criminal record?

20. Is it reasonable to use the binomial distribution to approximate the hypergeometric distribution in the following scenarios? (It depends, of course, on how accurate we need to be, and in most situations we prefer the exact answer to an approximation. Answer these questions based on the rough guideline given in the text.)

    (a) Suppose that 8 people are randomly selected from a room containing 10 men and 10 women. Is it reasonable to use the binomial distribution to approximate the probability that exactly 4 of the people selected are men?
    (b) Twelve balls are drawn without replacement from a large urn containing 100 white balls and 200 red balls. Is it reasonable to use the binomial distribution to approximate the probability that exactly 3 balls are white?
    (c) A card is drawn from a well-shuffled deck. The card is placed back in the deck, the cards are shuffled and another card is drawn. The process is repeated for a total of 60 draws. Is it reasonable to use the binomial distribution to calculate the probability that exactly 7 cards are kings?

5.7 The Poisson Distribution

5.7.1 Introduction

21. Suppose X is a Poisson random variable with parameter $\lambda = 5$.

    (a) What is $P(X = 2)$?
    (b) What is $P(X \leq 2)$?
    (c) What is $P(X > 2)$?
    (d) What is $P(2 < X \leq 5)$?
    (e) What is $P(X = 2|X \leq 3)$
    (f) What is the mean of X?
    (g) What is the standard deviation of X?

22. On a certain region of the Florida coast, shark attacks on humans occur at a rate of approximately 4 attacks per year. Assume that the number of attacks follows a Poisson distribution.

    (a) In a given year, what is the probability there is exactly one attack?
    (b) In a given year, what is the probability there are no more than 2 attacks?
    (c) What is the probability that in a given **two** year period there are exactly 4 attacks?

23. Which of the following statements are true?

    (a) A Poisson random variable can take on a countably infinite number of possible values.
    (b) A Poisson random variable can take on negative values.
    (c) A Poisson random variable represents a count of the number of occurrences of an event.
    (d) The mean and variance of a Poisson random variable are always equal.
    (e) If X has a Poisson distribution, then $P(X = 0) < P(X = 1)$.
    (f) Every random variable that represents a count is a Poisson random variable.

5.7.2 The Relationship Between the Poisson and Binomial Distributions

24. In which one of the following situations would the Poisson distribution provide the most reasonable approximation to the binomial distribution?

    (a) X has a binomial distribution with $n = 5$ and $p = 0.5$.

(b) X has a binomial distribution with $n = 10,000$ and $p = 0.9$.
(c) X has a binomial distribution with $n = 10$ and $p = 0.2$.
(d) X has a binomial distribution with $n = 500$ and $p = 0.01$.
(e) X has a binomial distribution with $n = 5$ and $p = 0.001$.

25. Protanopia is a kind of colour blindness that affects approximately 1% of males (and a much smaller percentage of females). Individuals with protanopia do not perceive red light normally, and have difficulty distinguishing between red and green and between red and blue. Suppose 100 males are randomly selected from a large population in which 1% of the males have protanopia.

    (a) Using the binomial probability mass function, find the probability that exactly 2 of these 100 males have protanopia. Also find the approximate probability based on the Poisson distribution.
    (b) Using the binomial probability mass function, find the probability that no more than 2 of these 100 males have protanopia. Also find the approximate probability based on the Poisson distribution.
    (c) Based on the binomial distribution, what is the variance of the number of males with protanopia? What is the variance based on the Poisson approximation?

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

26. Suppose it is known that the number of students entering a University Centre averages 4.5 per minute between noon and 12:30 pm. Would the number of students that enter this University Centre in a randomly selected minute in this time frame follows a Poisson distribution with $\lambda = 4.5$? Why or why not?

27. Suppose a very large pasture has a large number of cows, with an average rate of 1 cow per 100 m^2. Would the number of cows in a randomly selected 100 m^2 area of this pasture follow a Poisson distribution?

5.8 The Geometric Distribution

28. Suppose a patient at a hospital in Canada is in dire need of a plasma transfusion. The person has type AB blood, and people with type AB blood can receive plasma only from other individuals with type AB blood. The hospital is out of AB plasma, but they have a long list of possible donors of unknown blood type. The donors are Canadian, and approximately 3% of the Canadian population has type AB blood.

Suppose these potential donors can be thought of as a random sample from the Canadian population.

(a) If potential donors are tested for the AB blood type, what is the probability the first donor with type AB blood occurs on the fifth person tested?
(b) If potential donors are tested for the AB blood type, what is the probability the first donor with type AB blood occurs on or before the third person tested?
(c) If potential donors are tested for the AB blood type, what is the probability the first donor with type AB blood occurs after the 30th person tested?
(d) What is the mean number of potential donors that must be tested in order to find one with type AB blood?
(e) What is the standard deviation of the number of potential donors that must be tested in order to find one with type AB blood?

29. Tim Horton's coffee shop chain has long held a "Roll up the Rim" contest, where purchasers of a hot beverage can roll up the rim of the cup, possibly revealing that the cup is a prize winner. The probability of winning a prize on each cup changes slightly from year to year, but it is often approximately $\frac{1}{6}$. Suppose that while the contest is running, you often buy cups of coffee from Tim Horton's.

(a) What is the probability your first winning cup comes on the fourth cup that you purchase? (While the probability of winning changes a tiny amount from cup to cup, as cups are revealed as winners or non-winners, there are millions of cups so the changes in probability are minuscule. For the purposes of these questions, assume that the cups can be considered independent.)
(b) What is the probability your first win comes on or before the fourth cup?
(c) What is the probability your first win comes after the 20th cup?
(d) On average, how many cups will need to be purchased in order to get the first winner?

30. Which of the following statements are true?

(a) The mean of the geometric distribution is $\frac{1}{p}$.
(b) The most likely value of a geometric random variable is always 1.
(c) If $p > 0.5$, then $P(X = 2) < P(X = 3)$.
(d) The variance of the geometric distribution is always 1.
(e) A geometric random variable can take on a countably infinite number of possible values.
(f) If X has a geometric distribution, then $Y = \frac{1}{X}$ also has a geometric distribution.

5.9 The Negative Binomial Distribution

31. In his younger days, your author frequently practiced his golf game. While practicing around the greens, he would often play a game in which he had to hole a certain number of chips before allowing himself to leave. Suppose on a certain afternoon, he decided that he would allow himself to leave once he chipped-in 4 times. (He must hole out on a chip 4 times before leaving.) Suppose that his probability of chipping-in on any individual chip is approximately 0.05, and that the chips can be considered independent. (While the probability of chipping-in would vary a little from chip to chip, and the chips would not truly be independent, these assumptions provide a reasonable approximate model.)

    (a) What is the probability that the fourth chip-in occurs on the 20th attempt?
    (b) What is the probability that the fourth chip-in occurs on the 20th or 21st attempt?
    (c) What is the mean number of attempts required to chip-in 4 times?
    (d) What is the standard deviation of the number of attempts required to chip-in 4 times?
    (e) Challenge: What is the probability that it takes more than 100 attempts for him to chip-in 4 times?

32. Suppose a door-to-door salesman needs to make 3 more sales in order to reach his monthly sales goal. He knows from a large body of past experience that any time he knocks on a door he has approximately a 1.5% chance of making a sale. Assume that this is the correct probability of making a sale each time he knocks on a door, and that it is reasonable to assume independence between doors.

    (a) What is the probability that his 3rd sale occurs on the 50th door?
    (b) What is the probability that his 3rd sale occurs on the 200th door?
    (c) What is the mean number of doors he must approach in order to make his 3 sales?
    (d) What is the standard deviation of the number of doors he must approach in order to make his 3 sales?
    (e) Challenge: What is the probability he must knock on more than 500 doors in order to make his 3 sales?

5.10 The Multinomial Distribution

33. There are 3 main types of fingerprint pattern: arches, loops, and whorls. In fingerprint analysis and classification, each finger falls into one of these 3 categories.

(In more complicated classification systems there are various subcategories of these groups.) Suppose that the distribution of fingerprint pattern on the left thumb of adults in Libya follows the following distribution:[1]

Pattern	Arches	Loops	Whorls
Probability	0.12	0.49	0.39

Suppose 10 Libyan adults are randomly selected and the fingerprint pattern on the left thumb is identified.

(a) What is the probability that 2 have arches, 3 have loops, and 5 have whorls?
(b) What is the probability that 4 have arches, 4 have loops, and 2 have whorls?
(c) What is the probability that exactly 1 has arches?
(d) What is the expectation of the number that have arches?

34. A typical European roulette wheel has 18 red slots, 18 black slots, and 1 green slot. The wheel is spun and a ball lands in one of these equally likely slots. The distribution of the colour of the slot the ball lands in is:

Colour	Red	Black	Green
Probability	$\frac{18}{37}$	$\frac{18}{37}$	$\frac{1}{37}$

(a) In the next 20 spins of the wheel, what is the probability the ball lands in a red slot 8 times, a black slot 10 times, and the green slot 2 times?
(b) In the next 20 spins of the wheel, what is the probability the ball lands in a red slot 14 times, a black slot 3 times, and the green slot 3 times?
(c) In the next 20 spins of the wheel, what is the probability the ball lands in a red slot exactly 11 times?
(d) What is the mean number of times the ball lands in a red slot in the next 20 spins?

35. Suppose $X_1, X_2, \ldots, X_k$ have a multinomial distribution with parameters n and $p_1, p_2, \ldots, p_k$. Which of the following statements are true?

(a) X_2 has a binomial distribution with parameters n and p_2.
(b) $E(X_i) = np_i$.
(c) $Var(X_i) = np_i(1 - p_i)$.
(d) $\sum_{i=1}^{k} p_i = 1$.

[1] Based on a study by Fayrouz et al. (2012). Relation between fingerprints and different blood groups. *Journal of Forensic and Legal Medicine*, 19:18–21

5.11 Chapter Exercises

5.11.1 Basic Calculations

36. Consider the following discrete probability distribution:

x	-5	10	50	100
$p(x)$	$1c$	$1c$	$8c$	$10c$

Where c is a constant.

(a) What is the value of c?
(b) What is $P(X > 21)$?
(c) What is $P(X > 21|X < 60)$?

37. Consider the following discrete probability distribution for a random variable X. Note that the largest value of X and its corresponding probability are missing.

x	10	20	30	?
$p(x)$	0.1	0.2	0.3	?

If $E(X) = 60$, what is the missing value of X?

38. The probability distribution for a random variable is given by

x	10	11	12	13
$P(X = x)$	0.4	0.3	0.2	?

(a) What is the value of the missing probability?
(b) What is $P(X > 11)$?
(c) What is the mean of this distribution?
(d) What is the variance?

39. Consider again the discrete probability distribution given in Question 38. Suppose 5 values are randomly sampled from this distribution.

(a) What is the probability that exactly three 10s are selected?
(b) What is the probability that at least one 10 is selected?
(c) Values are sampled from this distribution repeatedly. What is the probability that the first 11 appears after the sixth sampled value?

40. Consider again the discrete probability distribution given in Question 38. Suppose 20 values are randomly sampled from this distribution.

(a) What is the probability that the number of 10s is less than 8?

(b) What is the probability that the number of 10s is between 2 and 6 (inclusive)?
(c) What is the expected number of 10s?

41. Let X represent the number of heads that come up when a fair coin is tossed 20 times. Let Y represent the number of fours that come up when a six-sided die is tossed 18 times. It is reasonable to assume that X and Y are independent.

    (a) What is the expected value of $X + Y$?
    (b) What is the standard deviation of $X - Y$?

42. Suppose a certain type of lottery ticket costs \$1. Let the random variable X represent the payout on a single ticket. The distribution of X is given in the following table.

x	0	1	10	100
$p(x)$	0.689	0.30	0.010	0.001

    (a) Find the probability of making a net profit (including the cost of the ticket).
    (b) If two tickets are purchased, and the tickets can be assumed to be independent, what is the probability the two tickets have a combined payout of \$100?
    (c) What is the expected value of the payout?
    (d) What is the standard deviation of the payout?
    (e) Given the payout is non-zero, what is the probability the payout is \$100?
    (f) What is the lottery's expected profit if 1000 of these tickets are sold?

5.11.2 Concepts

43. Show that $E[(X - \mu)^2] = E(X^2) - [E(X)]^2$. (You may use the property that $E(a + bX) = a + bE(X)$.)

44. Is the following statement true? Justify your response.

    Since $E(X^2) = E(XX) = E(X)E(X)$, X is independent of itself.

45. Are the following statements true or false?

    (a) If X and Y are two random variables that both have a mean of 0, X and Y must be independent.
    (b) If $E(X) > E(Y)$, then $Var(X) > Var(Y)$.
    (c) If $E(X) = 1$, then $E(X^2) = 1$

46. Suppose that X has a binomial distribution with $n = 15$ and $p = 0.2$, Y has a Poisson distribution with $\lambda = 6$, and Z has an unknown distribution with a mean of 5 and standard deviation of 18. Suppose that X, Y, and Z are all independent.

(a) What is $E(X+Y+Z)$?
(b) What is the standard deviation of $X+Y-Z$?
(c) Could the value of $X+Y-Z$ possibly be negative?

47. Fifteen children are let out during recess onto a playground that has an area of 1000 square metres. Each child detests every other child, and so they space themselves out as far as possible on the playground. Suppose a 1 square metre area of this playground is randomly selected. The number of children in the selected area is a random variable that has which one of the following distributions?

    (a) A binomial distribution with $n = 15$.
    (b) A Poisson distribution.
    (c) A geometric distribution
    (d) A continuous distribution.
    (e) None of the above.

48. Tom has a job of completing a delicate finishing procedure during the production of expensive pieces of high-tech equipment. If Tom makes a mistake, the equipment will be destroyed (at a high cost to the manufacturer). Tom is good at his job and on average has a probability of 0.001 of destroying an individual piece of equipment. But Tom is a bit shaky psychologically, and if he destroys a piece of equipment he gets rattled and is more likely to destroy other pieces he works with. Let X represent the number of pieces Tom destroys in the next 20 pieces he works on. Then X:

    (a) Has a binomial distribution with $n = 20$ and $p = 0.001$.
    (b) Has a Poisson distribution with $\lambda = 0.02$.
    (c) Is a discrete random variable, but does not have a binomial or Poisson distribution.
    (d) Is not a discrete random variable.
    (e) None of the above.

49. Which of the following statements are true?

    (a) A discrete random variable can take on a countable number of possible values.
    (b) A discrete random variable always represents a count.
    (c) The expected value of a discrete random variable must equal one of the variable's possible values.
    (d) The binomial distribution has two parameters: n and p.
    (e) A binomial random variable can take on one of $n + 1$ possible values.

50. Which of the following statements are true?

    (a) The variance and standard deviation of a discrete random variable cannot be

equal.
(b) The mean of a random variable cannot be negative.
(c) The standard deviation of a random variable cannot be negative.
(d) If X is a discrete random variable, then $Y = 2X + 3$ is a discrete random variable.

51. Which of the following statements are true?

(a) In a binomial setting, the probability of success usually changes from trial to trial.
(b) The mean of a Poisson random variable can be negative.
(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.
(d) If $p > 0$, then the variance of a binomial random variable is less than its mean.
(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.

5.11.3 Applications

52. The insect *Megamelus scutellaris* can feed on the sap of the water hyacinth, and it has been used as a method of biocontrol for this invasive plant species. *M. scutellaris* mates on the water hyacinth, and females create oviposition scars on the plant, laying one or more eggs in each scar. The approximate distribution of the number of eggs per scar is given in the following table. (Loosely based on a study by Sosa et al. (2005).)

x	1	2	3	4
$p(x)$	0.38	0.52	0.05	0.05

(a) What is the expected value of the number of eggs in a randomly selected oviposition scar? In other words, if we let X represent the number of eggs, what is $E(X)$?
(b) What is the standard deviation of the number of eggs?
(c) What is the expected value of the natural log of the number of eggs? That is, what is $E[\ln(X)]$?
(d) What is $E(X^2)$?

53. The Ontario Lottery Corporation runs a “Daily Keno” lottery in which 20 numbers are randomly selected without replacement from the integers 1 through 70. Before the drawing, keno players choose a set of numbers (anywhere from 2 to 10 numbers), and their ticket is a winner if enough of their numbers are randomly chosen. The

following table represents the possible payouts and their probabilities of occurring for a $1 six-number ticket.[2]

Number correct	4 or fewer	5	6
Payout	0	$25	$1000
Probability	0.9937921	0.0059123	0.0002956

(a) Verify the probabilities given in the table.
(b) What is the expectation of the payout on a $1 six-number ticket?
(c) What is the standard deviation of the payout on a $1 six-number ticket?
(d) For every $1 spent on this type of ticket, on average how much is the government keeping?

54. Consider again the information in Question 53. Suppose that you buy a single six-number keno ticket in each of the next 50 drawings.

(a) What is the probability that exactly one ticket wins $25?
(b) What is the probability that at least one ticket wins $25?
(c) What is the probability that at least one ticket wins a prize?
(d) What is the expected number of tickets that will win a prize?
(e) What is the standard deviation of the number that will win a prize?

55. You have an agreement with a cheese supplier, which states that he will sell you 1,000 kg of mozzarella cheese for $7000. The supplier is a bit of a gambler and on March 10 offers you an alternative arrangement: If it does not snow on April 10 you pay only $5,000, but if it snows that day you pay $100,000 instead. Suppose the probability of snowfall on April 10 is 0.02.

(a) If you choose the alternative arrangement, how much will you owe the supplier on average?
(b) Which deal is cheaper on average?
(c) Give at least one reason why it is not always wise to choose the option that results in the greatest expected value.

56. A manufacturer needs to ship an expensive part across the country. The part is very fragile, and if it breaks it will cost the manufacturer $25,000. A company offers insurance against breakage at a cost of $1,000.

(a) If the probability of breakage is 0.01, what is the expected profit for the insurance company?
(b) What is the standard deviation of the insurance company's payout?

[2] The listed payouts were extracted from http://www.olg.ca/lotteries/games/howtoplay.do?game=daily_keno#pt on January 23, 2015. The probabilities have been rounded to the 7th decimal place.

(c) What probability of breakage makes this a break-even proposition for the manufacturer?

57. People in need of a bone marrow transplant need to find a willing donor that is closely matched in human leukocyte antigens (HLA). The genes that govern HLA are located on a single chromosome (chromosome 6). Since chromosomes are inherited from a person's father and mother, any full sibling of a person has a $\frac{1}{4}$ chance of being an HLA-identical match.

    Suppose that a person in need of a bone marrow transplant has 6 siblings (none of whom are the result of a multiple birth).

    (a) What is the distribution of the number of their siblings that are an HLA-identical match?
    (b) What is the probability that at least one of their siblings is an HLA-identical match?
    (c) What is the probability that at least two of their siblings are an HLA-identical match?
    (d) What is the mean number of siblings that are an HLA-identical match?

58. For some types of surveys conducted by random digit dialling, if the call is answered by an adult at a residence, there is approximately a 9% chance that the person will complete the survey.[3] Suppose you work at a call centre that conducts a survey of this type.

    For the following questions, "call" refers to a call that is answered by an adult at a residence.

    (a) In the next 100 calls, what is the probability that exactly 8 people complete the survey?
    (b) In the next 100 calls, what is the probability that 8, 9, or 10 people complete the survey?
    (c) In the next 100 calls, what is the probability that at least 2 people complete the survey?
    (d) What is the mean number of surveys that will be completed in the next 100 calls ?
    (e) What is the standard deviation of the number of surveys that will be completed in the next 100 calls ?

59. Consider again the information in Question 58. Suppose you will get to leave for the day once you get the survey completed by 4 adults.

[3] Estimated from information on the Public Works and Government Services Canada website http://www.tpsgc-pwgsc.gc.ca/rop-por/rapports-reports/telephone/introduction-eng.html. Accessed January 2015.

For the following questions, "call" refers to a call that is answered by an adult at a residence.

(a) What is the probability that the 4th completed survey occurs on the 50th call?
(b) What is the probability that the 4th completed survey occurs on 50th or 51st call?
(c) What is the mean number of calls needed to get 4 completed surveys?
(d) What is the standard deviation of the number of calls needed to get 4 completed surveys?
(e) Challenge: What is the probability it will take more than 100 calls to get the 4 completed surveys?

60. A 2010 investigation by *Consumer Reports* studied fresh whole broiler chicken at U.S. retailers. Among other things, they found that salmonella was detectable in 14% of chickens. For the purposes of this question, assume that this figure is correct, and 14% of chickens available at retailers have detectable levels of salmonella contamination.

    (a) If 12 chickens are randomly selected from U.S. retailers, what is the probability that exactly 2 have detectable levels of salmonella?
    (b) If 12 chickens are randomly selected from U.S. retailers, what is the probability that no more than 2 have detectable levels of salmonella?
    (c) Suppose a retailer claims that on average, no more than 2% of their chickens have salmonella. In a random sample of 20 chickens from this retailer, 3 chickens are found to have salmonella contamination. If the retailer's claim is true, what is the probability of seeing at least 3 chickens with salmonella contamination in a sample of 20 chickens?

61. Among eligible blood donors in Canada, approximately 10% have given blood at some point. Suppose we draw a random sample of 5 eligible blood donors in Canada.

    (a) What is the distribution of the number that have given blood at some point?
    (b) What is the probability exactly one has given blood at some point?
    (c) What is the probability that at least 2 have given blood at some point?
    (d) What is the probability that no more than 2 have given blood at some point?
    (e) If *at least* one of these five people has given blood at some point, what is the probability that *exactly* one of these 5 has given blood at some point?

62. You have been giving a friend of yours \$1 a week for the past three years (156 weeks) to buy lottery tickets. Each week, your friend has supposedly been contributing \$1, buying a \$2 Lotto 6/49 ticket, checking the numbers, and sharing the wins and losses with you. You have not won a cash prize in 156 weeks, and you are growing suspicious of your friend. (According to the Ontario Lottery and Gaming Corporation, the probability of winning a cash prize in a single Lotto 6/49 drawing

is approximately 0.03.)

If one ticket is purchased per week for 156 weeks:

(a) What is the distribution of the number of tickets that win a cash prize?
(b) What is the expected number of tickets that win a cash prize?
(c) What is the probability that no ticket wins a cash prize?
(d) Is there strong evidence that your friend has been ripping you off?

63. Suppose that the number of fatal crashes per year involving U.S. commercial airlines has a mean of approximately 1.1.[4] To a reasonable approximation, fatal crashes can be thought of as occurring randomly and independently.

(a) What is the probability that there are no fatal crashes involving U.S. commercial airlines in the next year?
(b) What is the probability that there are 2 or 3 fatal crashes involving U.S. commercial airlines in the next year?
(c) What is the probability that there are more than 4 fatal crashes involving U.S. commercial airlines in the next year?
(d) What is the standard deviation of the number of fatal crashes involving U.S. commercial airlines in the next year?

64. Consider again the information in Question 63.

(a) What is the probability that there is at least 1 fatal crash involving U.S. commercial airlines in the next 2 year period?
(b) What is the probability that there is more than 1 fatal crash involving U.S. commercial airlines in the next 2 year period?
(c) What is the mean number of fatal crashes involving U.S. commercial airlines in the next 2 year period?
(d) What is the standard deviation of the number of fatal crashes involving U.S. commercial airlines in the next 2 year period?

65. According to Wikipedia, objects with a diameter in excess of 1 km hit the earth at a rate of approximately two every million years. Assume that the number of impacts in any given time period has (approximately) a Poisson distribution.

(a) What is the approximate probability that there is at least one impact of this size in a given 100,000 year period?

[4] This value is the mean number of annual fatal crashes for the period 1990-2012, calculated from data obtained from the National Transportation Safety Board (http://www.ntsb.gov/investigations/data/Pages/paxfatal.aspx). It is likely not very far from the true value. Let's assume it to be correct for the purposes of this question.

(b) What is the approximate probability that there is at least one impact of this size in a given 10,000 year period?

66. In a famous series of genetics experiments, Gregor Mendel investigated inheritance in sweet peas. In one experiment, he crossed one pure line of peas that had round yellow seeds with another pure line that had wrinkled green peas. A round shape and yellow colour are dominant traits, so the peas resulting from the first generation cross all had round yellow seeds (they had the round/yellow *phenotype*). This first generation was then self-crossed. For the second generation, under Mendelian inheritance with independent assortment, a 9:3:3:1 ratio of phenotypes would be expected. In other words, the distribution of phenotypes would be:

Phenotype	Round/Yellow	Round/Green	Wrinkled/Yellow	Wrinkled/Green
Probability	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

Suppose we examine 50 pea plants from this type of cross.

(a) What is the distribution of the number of round/yellow plants?
(b) What is the mean number of round/yellow plants?
(c) What is the standard deviation of the number of round/yellow plants?
(d) What is the probability that there are exactly 4 wrinkled/green plants?
(e) What is the probability that there are 28 round/yellow plants, 12 round/green plants, 6 wrinkled/yellow plants, and 4 wrinkled/green plants?

5.11.4 Extra Practice Questions

67. Suppose that for a binomial random variable X, $n = 10$ and $\mu = 2.4$.

(a) What is p?
(b) What is $P(X = 2)$?
(c) What is σ?

68. A certain coin sold by a novelty store has a probability of coming up heads of $\frac{2}{3}$. Let X be a random variable representing the total number of heads when this coin is tossed twice.

(a) What is the mean of X?
(b) What is the standard deviation of X?

69. According to the *Pocket World in Figures* by the *Economist*, 17.8% of the Canadian population is over 60 years old. Suppose 20 Canadians are randomly selected.

(a) What is the distribution of the number of people in the sample that are over 60 years old?
(b) What is the probability that at least 2 are over 60 years old?
(c) What is the probability that exactly 2 are over 60 years old?
(d) What is the probability that none are over 60 years old?
(e) What is the expected number of people that are over 60 years old?

70. Approximately 60% of Canadian adults voted in the last federal election. Suppose we draw a random sample of five people who were Canadian adults at the time of the last election.

(a) What is the distribution of the number of people in the sample that voted in the last federal election?
(b) What is the probability that exactly 4 voted in the last federal election?
(c) What is the probability that four or more voted in the last federal election?
(d) Given at least four voted, what is the conditional probability that all 5 voted?
(e) What is the mean of the number that voted in the last election?

71. A 2011 report by Statistics Canadashowed the unemployment rate in Canada was approximately 7.6% (7.6% of the labour force was out of work). If 14 members of the Canadian labour force were randomly selected during this time period:

(a) What is the probability exactly three were out of work?
(b) What is the probability more than 2 were out of work?
(c) Given at least 2 of the 14 were out of work, what is the probability at least 1 of the 14 was out of work?
(d) Given at least 3 of the 14 were out of work, what is the probability no more than one of the 14 was out of work?

72. Approximately 35% of cars that are at least 13 years old fail the Ontario Drive Clean emissions test. Suppose we draw a random sample of 10 emissions tests on cars that are at least 13 years old.

(a) What is the probability that exactly 2 fail?
(b) What is the probability that no more than 2 fail?
(c) What is the probability that more than 2 fail?
(d) What is the expectation of the number of failures?
(e) What is the standard deviation of the number of failures?
(f) Suppose a certain garage has given a failing emissions grade to each of the 10 cars they have tested that are at least 13 years old. Is this strong evidence the garage is fraudulent?

73. A certain amount of a material is experiencing radioactive decay, with a decay

rate of 0.8 per second (an average of 0.8 radioactive decays per second). It is reasonable to assume the number of decays in a given time period will follow a Poisson distribution.

(a) In a one second period, what is the probability the material has exactly two radioactive decays?
(b) In a three second period, what is the probability the material has exactly two radioactive decays?
(c) In a three second period, what is the probability the material has no more than two radioactive decays?

74. At a wood chipper in a paper mill, log jams occur (approximately) randomly and independently, at the rate of 0.2 per hour.

(a) What is the probability that there are exactly 3 log jams in an 8 hour day?
(b) What is the probability that there is at least one log jam in an 8 hour day?
(c) Challenge: In 12 randomly selected hours, what is the probability that there is at least one log jam in at least 4 of the hours?.

75. Airplanes arrive at a small airport at an average rate of 5 per hour. Suppose the number of planes arriving in any given period approximately follows a Poisson distribution.

(a) What is the probability that in a 2 hour period, at least one plane arrives?
(b) What is the probability that in a 3 hour period, exactly 3 arrive?
(c) What is the probability that in a half-hour period, no planes arrive?
(d) What is the standard deviation of the number of planes that arrive in an 8 hour period?

76. Suppose that planes at a remote airport are very poorly maintained, and have a 0.01 probability of crash landing before their destination. Assuming independence, in the next 10 planes that take off:

(a) What is the distribution of the number that crash?
(b) What is the probability that at least one crashes?
(c) What is the probability that exactly one crashes?
(d) What is the probability they all crash?
(e) What is the standard deviation of the number that crash?

77. The probability that a randomly selected Canadian is a medical doctor is approximately 0.002.

(a) If Canadians are randomly selected, what is the probability the fourth person selected is the first that is a medical doctor?

(b) What is the probability that more than 1000 Canadians must sampled before encountering the first medical doctor?
(c) If 80 Canadians are randomly selected, what is the probability that exactly one is a medical doctor?
(d) If 80 Canadians are randomly selected, what is the probability that more than two are medical doctors?

78. A company is interested in purchasing a new automatic welding machine. The producer of the machine claims that for a certain type of weld, each weld has a 90% chance of being successful.

    Suppose that the company interested in the welder decides to test it by having it perform welds until it fails for the first time. Assume that the producer's claim is true (each weld will be successful with probability 0.90 and unsuccessful with probability 0.10), and that the trials are independent.

    (a) What is the distribution of the number of welds needed to get the first unsuccessful weld?
    (b) What is the probability that the first unsuccessful weld occurs on the third trial?
    (c) What is the probability that the first unsuccessful weld occurs on or before the third trial?
    (d) What is the probability that the first unsuccessful weld occurs after the 13th trial?

79. Consider again the information in Question 78. Suppose the company purchases the machine and uses it whenever a weld of the appropriate type is required. (Assume that the probability of a successful weld is always 0.90 and that the trials are independent.)

    (a) What is the probability that the third unsuccessful weld occurs on the 22nd welding job?
    (b) In the next 20 welds, what is the probability that exactly 3 are unsuccessful?
    (c) In the next 20 welds, what is the probability that at least 3 are unsuccessful?
    (d) What is the mean number of unsuccessful welds in a sample of 20 welds?

80. Suppose that you get a stimulating job inspecting shipments of light bulbs that arrive at a factory. Your job is to take out one randomly selected bulb and test it. If the bulb works, the shipment is accepted. If the bulb malfunctions then the shipment is rejected. Unknown to you, in reality the probability that any randomly selected bulb malfunctions is 0.04, and the shipments can be considered to be independent.

    (a) What is the probability that the first shipment you must reject occurs on the

10th shipment?

(b) What is the probability that the first shipment you must reject occurs after the 10th shipment?

(c) In the first 60 shipments, what is the probability you must send back no more than one shipment?

81. A room contains 30 women and 55 men, and 5 of these people are randomly selected without replacement,

(a) What is the exact probability of getting exactly 4 men in the sample?

(b) Use the binomial distribution to approximate the probability of getting 4 men in the sample.

82. Suppose the distribution of dandelions in a large meadow is approximately Poisson, with a mean of 4 dandelions per square metre. A 1 m^2 area in this meadow is randomly selected.

(a) What is the probability there are exactly 2 dandelions in the area?

(b) What is the probability there are more than 2 dandelions in the area?

(c) Given there is at least one dandelion in the area, what is the probability there are exactly three dandelions?

(d) What is the standard deviation of the number of dandelions in the randomly selected area?

83. Suppose you design a new type of dental implant. You feel that if the implant is placed in a randomly selected dental patient, the probability that it does not take properly is 0.05. As part of a test of the new implant, you intend to place your implant into patients until you get 3 unsuccessful implants. (Assume for this question that your probability assessment is correct, and also assume that the patients can be considered independent.)

(a) What is the probability that the first unsuccessful implant occurs on the fourth trial?

(b) What is the probability that the first unsuccessful implant occurs after the 11th trial?

(c) What is the probability that the third unsuccessful implant occurs on the 28th trial?

84. A dollar store bin contains 15 rubber footballs, 20 rubber soccer balls, and 8 rubber volleyballs. Suppose 6 of these balls are randomly selected without replacement.

(a) What is the probability that exactly 2 footballs are selected?

(b) What is the expectation of the number of footballs selected?

(c) What is the probability that no more than 1 soccer ball is selected?

(d) What is the probability 3 footballs, 2 soccer balls, and 1 volleyball are selected?

85. A room contains 5 biomedical engineers, 7 computer engineers, 5 mechanical engineers, and 3 water resource engineers.

    (a) If 6 people are randomly selected without replacement, what is the probability there are exactly 3 computer engineers selected in the group of 6?
    (b) If 6 people are randomly selected without replacement, what is the probability there are exactly 2 biomedical engineers, 2 computer engineers, 1 mechanical engineer, and 1 water resource engineer in the sample?

86. A rental car company has 20 cars available: 4 Hondas, 6 Fords, 3 Toyotas, and 7 Chrysler vehicles. If they randomly pick 5 of these cars for inspection (without replacement), what is the probability the sample consists of 1 Honda, 1 Ford, 1 Toyota, and 2 Chrysler vehicles?

87. Consider the following distribution of a categorical variable.

Category	A	B	C	D
Probability	0.4	0.3	0.2	0.1

Suppose 5 values are sampled independently from this distribution. What is the probability there will be three A's, one B, and one C?

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. Test your conceptual understanding: Which of the following statements are true? (You should be able to explain *why* a statement is true or false.)

   (a) For continuous random variables, probabilities correspond to areas under the density curve.
   (b) The area under any continuous probability distribution is 1.
   (c) A probability density function can take on negative values.
   (d) A probability density function can take on values greater than 1.
   (e) $f(x) = P(X = x)$ for all x.
   (f) $f(x) = \mu$ whenever the distribution is symmetric.

6.2.1 An Example Involving Integration

2. Warning: This question requires calculus (a basic integration).

   Suppose the random variable X has the probability density function:

$$f(x) = \begin{cases} cx^2 & \text{for } 0 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) What value of c makes this a legitimate probability distribution?
   (b) What is $P(X > 2)$?
   (c) What is $P(X < 1)$?
   (d) What is the 32nd percentile of this distribution?
   (e) What is the median of this distribution?
   (f) What is the mean of this distribution?

3. Warning: This question requires calculus (a basic integration).
   Let the random variable X have the probability density function:

$$f(x) = \begin{cases} \dfrac{c}{x^2} & \text{for } 4 < x < 10 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) What value of c makes this a legitimate pdf?
   (b) What is $P(8 < X < 9)$?
   (c) What is the 44th percentile of this distribution?

4. Warning: This question requires calculus—a basic integration.

   Suppose the response time of a certain computer system approximately follows an exponential distribution with a mean of 0.10 seconds:

$$f(x) = 10e^{-10x}$$

   for $x > 0$.

   (a) What is the probability a randomly selected response time exceeds 0.22 seconds?
   (b) What is the probability a randomly selected response time exceeds 0.02 seconds?
   (c) What is the probability a randomly selected response time lies between 0.08 and 0.12 seconds?
   (d) What is the 25th percentile of response time for this computer system?
   (e) What is the median response time for this computer system?

6.3 The Continuous Uniform Distribution

5. Let X have a continuous uniform distribution on the interval [5, 25].

   (a) What is $f(x)$?
   (b) What is $P(X = 8.0)$?
   (c) What is $P(X > 8.0)$?
   (d) What is $P(X > 8.0|X < 7.3)$?
   (e) What is $P(X > 8.0|X > 7.3)$?
   (f) What is the median of this distribution?
   (g) What is the mean of this distribution?
   (h) What is the 47^{th} percentile of this distribution?
   (i) What is the interquartile range of this distribution?

6. Which of the following statements about the continuous uniform distribution are true? (You should be able to explain *why* a statement is true or false.)

   (a) The uniform distribution is symmetric.
   (b) For any uniform distribution, the mean and median are equal.
   (c) For any uniform distribution, $Q_1 = -Q_3$.
   (d) For any uniform distribution, $Q_3 - Q_2 = Q_2 - Q_1$.
   (e) If a random variable has a uniform distribution, then it cannot take on negative values.
   (f) If a random variable has a uniform distribution, then its standard deviation is greater than its mean.

6.4 The Normal Distribution

7. What are the parameters of the normal distribution? Can any of these parameters be negative?

8. Which of the following statements are true?

   (a) The normal distribution is symmetric about μ.
   (b) If X is a normally distributed random variable, then $P(X < 0) > 0$.
   (c) If X is a normally distributed random variable, then $P(X = 0) = 0$.
   (d) The standard deviation of a normal distribution cannot be greater than the mean.
   (e) The mean and median of a normal distribution are always equal.

6.4.1 Finding Areas Under the Standard Normal Curve

9. Suppose the random variable Z has the standard normal distribution.

   (a) What is $P(Z < 0.37)$?
   (b) What is $P(Z > 0.37)$?
   (c) What is $P(0 < Z < 0.37)$?
   (d) What is $P(Z < -1.28)$?
   (e) What is $P(Z > -1.28)$?
   (f) What is $P(Z = -1.28)$?
   (g) What is $P(-1.28 < Z < 0)$?
   (h) What is $P(Z > 2.41|Z > 0)$?
   (i) What is $P(-2.41 < Z < 2.41|Z > 0)$?
   (j) What is the 83^{rd} percentile of the standard normal distribution?
   (k) What is the interquartile range of the standard normal distribution?
   (l) If two values are independently sampled from the standard normal distribution, what is the probability they are both less than 2.13?

10. Suppose the random variable Z has the standard normal distribution.

    (a) What is $P(Z > -2.73)$?
    (b) What is $P(Z > 2.73)$?
    (c) What is the value z such that $P(Z > z) = 0.1210$?
    (d) What is the value z such that $P(Z < z) = 0.1210$?
    (e) What is the value z such that $P(-z < Z < z) = 0.95$?
    (f) What is the value of z such that $P(Z > z) = 0.7611$?
    (g) If two values are sampled independently from the standard normal distribution, what is the probability that both values are greater than 2.00?

11. $P(Z > 3178890557384735892750948 7)$ is closest to which one of the following?

    (a) 0
    (b) 0.5
    (c) 1
    (d) 31788905573847358927509487
    (e) ∞

6.4.2 Standardizing Normally Distributed Random Variables

12. Suppose the random variable X has a normal distribution with a mean of $\mu = 120$ and a standard deviation of $\sigma = 20$.

(a) What is $P(X < 105)$?
(b) What is $P(X > 87)$?
(c) What is $P(92 < X < 108)$?
(d) What is $P(82 < X < 87)$?
(e) What is $P(X = 84)$?
(f) What is the 20th percentile of the distribution of X?
(g) What is the 80th percentile of the distribution of X?
(h) What is the interquartile range of the distribution of X?

13. A study[1] found that birth weights of female African elephants born in captivity are approximately normally distributed with a mean of 95.1 kg and a standard deviation of 13.7 kg. (The following questions are all in reference to African elephants born in captivity. The distribution of weights of elephants born in the wild is likely different from that of those born in captivity.)

(a) What is the probability that a randomly selected female newborn African elephant weighs more than 120 kg?
(b) What is the probability that a randomly selected female newborn African elephant weighs less than 100 kg?
(c) What is the probability that a randomly selected female newborn African elephant weighs between 90.0 and 110.0 kg?
(d) What is the 30th percentile of birth weights of female African elephants?
(e) What is the 80th percentile of birth weights of female African elephants?

14. Test your conceptual understanding: Which of the following statements are true? (You should be able to explain *why* a statement is true or false.)

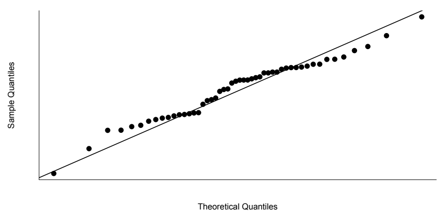
(a) If X has a normal distribution with mean μ and standard deviation σ, then $\frac{X-\mu}{\sigma}$ has the standard normal distribution.
(b) If Z has the standard normal distribution, then the mean and standard deviation of Z are equal.
(c) The first quartile of the standard normal distribution is greater than 0.
(d) The first quartile of any normal distribution is greater than 0.
(e) If X has a normal distribution, then $P(X > \mu) = P(X < \mu)$.

[1]Dale, R. (2010). Birth statistics for African (*Loxodonta africana*) and Asian (*Elephas maximus*) elephants in human care: History and implications for elephant welfare. *Zoo Biology*, 29:87–103. (The mean and standard deviation given in this question are estimates based on this article, but assume that they are the true values for the purposes of this question.)

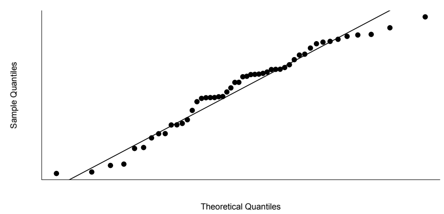
6.5 Normal Quantile-Quantile Plots: Is the Data Approximately Normality Distributed?

6.5.1 Examples of Normal QQ Plots for Different Distributions

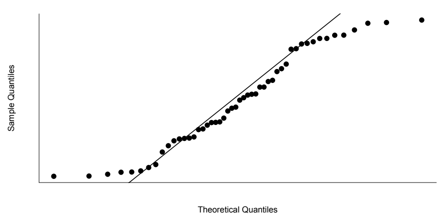
15. Consider the plots Figure 6.1, which represent normal quantile-quantile plots for 4 different samples.



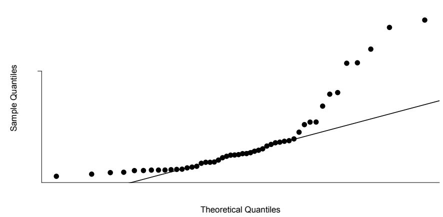
(a) A



(b) B



(c) C



(d) D

Figure 6.1: Four normal QQ plots.

   (a) What does Plot A tell us about the shape of the distribution from which this sample is drawn?
   (b) What does Plot B tell us about the shape of the distribution from which this sample is drawn?
   (c) What does Plot C tell us about the shape of the distribution from which this sample is drawn?
   (d) What does Plot D tell us about the shape of the distribution from which this sample is drawn?

16. Draw a sketch of a normal quantile-quantile plot in which the data shows strong left skewness, but has one large outlier (one value that is much larger than would be expected under normality).

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. Suppose random variable X has the uniform distribution on the interval $[0, 50]$.

    (a) What is $P(X < 10)$?
    (b) What is $P(X > 10)$?
    (c) What is $P(12 < X < 23)$?
    (d) What is $P(-5 < X < 5)$?
    (e) What is the value a such that $P(X > a) = 0.80$?
    (f) What is the 20th percentile of this distribution?
    (g) What is the median of this distribution?

18. Consider the continuous probability distribution illustrated in Figure 6.2.

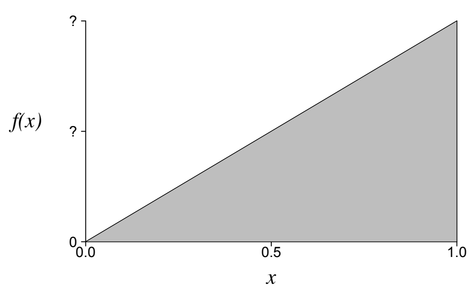


Figure 6.2: A continuous probability distribution.

    (a) What is the height of the curve at $x = 1$?
    (b) What is the height of the curve at $x = 0.50$?
    (c) What is the probability density function?

(d) What is the height of the curve at $x = 2$?
(e) What is $P(X < 1.5)$?
(f) What is $P(X < 0.5)$?
(g) What is $P(X > 0.5)$?
(h) What is the median of this distribution?

6.7.2 Concepts

19. $P(Z > -458958643961346)$ is closest to which one of the following?

    (a) 0
    (b) 0.5
    (c) 1
    (d) -458958643961346
    (e) ∞

20. Test your conceptual understanding: Which of the following statements are true? (You should be able to explain *why* a statement is true or false.)

    (a) The 95th percentile of a continuous random variable cannot be less than 0.
    (b) The mean of a continuous random variable cannot be negative.
    (c) The mean of a continuous random variable can be greater than the third quartile.
    (d) The normal distribution is symmetric.
    (e) If X has a normal distribution, then the mean and median of X are equal.

6.7.3 Applications

21. A study found that the total serum cholesterol level in Canadian men aged 60–79 is approximately normally distributed with a mean of 4.7 mmol/L and a standard deviation of 1.2 mmol/L.[2]

    (a) High total serum cholesterol levels have been linked to heart disease. By one standard, a total serum cholesterol level of at least 6.2 mmol/L is considered high. If one Canadian man in this age group is randomly selected, what is the probability his total serum cholesterol level is at least 6.2 mmol/L?

[2] The mean and standard deviation are estimates based on information from the Canadian Health Measures Survey (Statistics Canada, 2013). Assume these are the true values for the purposes of this question.

(b) What proportion of Canadian men in this age group have high total serum cholesterol? (As judged by the 6.2 mmol/L guideline.)
(c) The 6.2 mmol/L cutoff value for high total serum cholesterol is quite high, and some sources suggest that a person should aim to keep their total serum cholesterol below 4.6 mmol/L. What proportion of Canadian males between 60 and 79 years of age have a total serum cholesterol level that is below 4.6 mmol/L?
(d) What proportion of Canadian males between 60 and 79 years of age have total serum cholesterol levels between 4.6 mmol/L and 6.2 mmol/L?
(e) What is the 20th percentile of total serum cholesterol levels for Canadian men in this age group?
(f) What is the 80th percentile of total serum cholesterol levels for Canadian men in this age group?

22. A study found that heights of Canadian females aged 20 to 39 are approximately normally distributed with a mean of 163.3 cm and a standard deviation of 6.4 cm. (See footnote 2.)

(a) What is the probability a randomly selected Canadian female in this age group is at least 6 feet (182.88 cm) tall?
(b) What is the probability a randomly selected Canadian female in this age group is taller than 170.0 cm?
(c) What is the probability a randomly selected Canadian female in this age group is shorter than 170.0 cm?
(d) What is the probability a randomly selected Canadian female in this age group is taller than 150.0 cm?
(e) What is the probability a randomly selected Canadian female in this age group has a height that is within 1 standard deviation of the mean height?
(f) What is the 20th percentile of heights of adult Canadian females in this age group?

23. A study found that heights of Canadian males aged 20 to 39 are approximately normally distributed with a mean of 177.7 cm and a standard deviation of 6.5 cm. (See footnote 2.)

(a) What is the probability a randomly selected Canadian male in this age group is at least 6 feet (182.88 cm) tall?
(b) What is the probability a randomly selected Canadian male in this age group is taller than 170.0 cm?
(c) What is the probability a randomly selected Canadian male in this age group is shorter than 170.0 cm?
(d) What is the 99th percentile of heights of Canadian males in this age group?
(e) If 3 Canadian males in this age group are randomly selected, what is the

probability they are all taller than 170.0 cm?

(f) If 2 Canadian males in this age group are randomly selected, what is the probability they are both at least 6 feet (182.88 cm) tall?

(g) If 5 Canadian males in this age group are randomly selected, what is the probability that exactly 2 of them are at least 6 feet (182.88 cm) tall?

24. Sea urchin gonads are a popular seafood in Japan. A type of green sea urchin (*Hemicentrotus pulcherrimus*) is harvested off the coast of Japan, but those harvested off the northeast coast are often found to have a bitter taste and are not commercially viable. A study[3] investigated characteristics of green sea urchins harvested off the northeast coast. In one part of the study, it was found that the test diameters of mature male green sea urchins in this area are approximately normally distributed with a mean of 37.9 mm and a standard deviation of 3.8 mm. (The test is the sea urchin's shell.)

    (a) What is the probability that a randomly selected mature male green sea urchin in this area has a test diameter that is greater than 40.0 mm?
    (b) What proportion of mature male green sea urchins in this area have a test diameter that is less than 30.0 mm?
    (c) What proportion of mature male green sea urchins in this area have test diameters that are between 30.0 mm and 40 mm?
    (d) What is the median test diameter of mature male green sea urchins in this area?
    (e) What is the 25th percentile of test diameters of mature male green sea urchins in this area?

25. A study[4] investigated the weight of cereal in bags of cereal purchased at a grocery store. A certain brand of cereal had a size of bag that had a nominal (stated) weight of 368 grams. It was found that the actual weight of cereal in these bags was approximately normally distributed with a mean of 374.5 grams and a standard deviation of 2.8 grams.

    (a) What proportion of bags of this type contain at least 380 grams of cereal?
    (b) What proportion of bags of this type contain less than the stated amount of cereal?
    (c) A bag of this type is randomly selected. What is the probability that it contains between 370 and 380 grams of cereal?
    (d) What is the 50th percentile of the weight of cereal in bags of this type?

[3] Murata, Y. (2001). *Studies on a novel bitter amino acid, pulcherrimine in the green sea urchin gonads.* PhD thesis, Kyoto University. (The mean and standard deviation in this question are estimates based on this article, but assume that they are the true values for the purposes of this question.)

[4] Personal study conducted by your author. (The mean and standard deviation in this question are estimates based on this study, but assume that they are the true values for the purposes of this question.)

(e) What is the 15th percentile of the weight of cereal in bags of this type?

26. A study[5] investigated the amount of protein in a type of 2% cow's milk. The amount of protein in a cup (245 grams) of milk was found to be approximately normally distributed with a mean of approximately 8.0 grams and a standard deviation of approximately 0.30 grams.

    (a) What is the probability that a randomly selected cup of this type of milk contains less than 8.10 grams of protein?
    (b) What is the probability that a randomly selected cup of this type of milk contains more than 14 grams of protein?
    (c) What is the probability that a randomly selected cup of 2% milk contains between 8.05 and 8.20 grams of protein?
    (d) What is the 99th percentile of the amount of protein in a cup of 2% milk?
    (e) What is the 75th percentile of the amount of protein in a cup of 2% milk?

27. The total amount of fat in one serving (85 grams) of pan-browned crumbled 90% lean ground beef is approximately normally distributed with a mean of 10.2 grams and a standard deviation of 1.8 grams. (See footnote 5.)

    (a) What is the probability that a randomly selected serving of this type of beef contains less than 15.0 grams of fat?
    (b) What is the probability that a randomly selected serving of this type of beef contains between 8.0 and 12.0 grams of fat?
    (c) If two servings of this type of beef are randomly selected, what is the probability that each serving contains more than 14.0 grams of fat?
    (d) What is the 15th percentile of the amount of fat in a serving of this type of beef?
    (e) What is the 95th percentile of the amount of fat in a serving of this type of beef?

28. Suppose that the head lengths of adult female lizards of the species *Phrynocephalus versicolor* are approximately normal with a mean of 11.90 mm and a standard deviation of 0.41 mm.[6]

    (a) If an adult female lizard of this species is randomly selected, what is the probability its head length is between 11.0 and 12.0 mm?
    (b) If an adult female lizard of this species is randomly selected, what is the probability its head length is at least 11.5 mm?

[5] The mean and standard deviation in this question are estimates based on information from the USDA National Nutrient Database. Assume that they are the true values for the purposes of this question.)

[6] The mean and standard deviation are based on a study by Qu et al. (2011). Assume they are the true values for the purposes of this question.

(c) A person captures a female lizard that they believe to be of this species. The lizard's head length is found to be 22.5 mm. Should they rethink their belief that this lizard is of this species?
(d) What is the 10th percentile of the head lengths for adult female *P. versicolor* lizards?
(e) What is the 90th percentile of the head lengths for adult female *P. versicolor* lizards?

29. Many studies have investigated the weight of a human organs, often basing the findings on post-mortem results from accidental deaths. In one study[7], it was found that the human heart in adult Caucasian males had a mean of 365.0 grams and a standard deviation of 71.0 grams. For the purposes of this question, assume that the heart weight of adult Caucasian males is approximately normally distributed with a mean of 365.0 grams and a standard deviation of 71.0 grams.

(a) What is the probability a randomly selected adult Caucasian male has a heart that weighs more than 400.0 grams?
(b) What is the probability a randomly selected adult Caucasian male has a heart that weighs more than a pound (453.6 grams)?
(c) What is the probability a randomly selected adult Caucasian male has a heart that weighs less than 300 grams?
(d) What is the probability a randomly selected adult Caucasian male has a heart that weighs between 300 and 400 grams?
(e) What is the 90th percentile of heart weight for adult Caucasian males?

30. Weights of full-term newborn babies in the United States are (very roughly) normally distributed with a mean of approximately 3.3 kg and a standard deviation of approximately 0.5 kg. (This is based on information found in a study by Martin et al. (2015).)

(a) If one full-term newborn baby is randomly selected, what is the probability it weighs less than 2.0 kg?
(b) If one full-term newborn baby is randomly selected, what is the probability it weighs less than 4.0 kg?
(c) What proportion of full-term newborn babies in the United States weigh between 2.0 and 4.0 kg?
(d) If one full-term newborn baby is randomly selected, what is the probability it weighs more than 6.0 kg?
(e) What is the 90th percentile of full-term newborn baby weights in the United States?

[7] Lorin de la Grandmaison et al. (2001). Organ weight in 684 adult autopsies: new tables for a Caucasoid population. *Forensic Science International*, 119:149–154.

(f) What is the 25th percentile of full-term newborn baby weights in the United States?

6.7.4 Extra Practice Questions

31. Under certain conditions, the lifetimes of a certain type of insect are approximately normally distributed with a mean of 60.2 days and a standard deviation of 18.0 days.

    (a) If one of these insects is randomly selected, what is the probability it lives longer than 100 days?
    (b) If one of these insects is randomly selected, what is the probability it lives longer than 200 days?
    (c) If one of these insects is randomly selected, what is the probability it dies before 50 days?
    (d) What is the probability a randomly selected insect dies between 50 and 70 days?
    (e) What is the 90th percentile of the lifetimes of these insects under these conditions?

32. A supplier of pure oxygen claims that their containers of pure oxygen have a mean purity of 99.500%, with a standard deviation of 0.048%. In your first dealing with this company, your first container has a purity level of 99.43%. Assume the purity levels are approximately normally distributed.

    (a) Assuming the company's claim is true, what is the probability that a randomly selected container has a purity level less than 99.43%?
    (b) Assuming the company's claim is true, what is the probability that a randomly selected container has a purity level greater than 99.60%?
    (c) Assuming the company's claim is true, what is the 10th percentile of purity levels?
    (d) Assuming the company's claim is true, what is the interquartile range (IQR) of purity levels?
    (e) Assuming the company's claim is true, what is the probability that your next 4 containers all have purity levels greater than 99.40%? (Suppose that it is reasonable to assume the containers are independent.)

33. The weights of a type of adult pigeon in a city are approximately normally distributed with a mean of 370 grams and a standard deviation of 28 grams.

    (a) If one adult pigeon is randomly selected, what is the probability it weighs less than 350 grams?

(b) If one adult pigeon is randomly selected, what is the probability it weighs more than 1 pound (453.6 grams)?
(c) If one adult pigeon is randomly selected, what is the probability it weighs between 350 and 400 grams?
(d) If one adult pigeon is randomly selected, what is the probability it weighs more than 6.0 kilograms?
(e) What is the 5th percentile of adult pigeon weights in this city?

34. At a certain large pizza restaurant, the daily demand for cheese is approximately normally distributed with a mean of 50 kg and a variance of 36 kg^2.

    (a) On a randomly selected day, what is the probability the restaurant needs more than 60 kg of cheese?
    (b) On a randomly selected day, what is the probability the restaurant needs less than 45 kg of cheese?
    (c) What is the 80th percentile of cheese demand?
    (d) If the restaurant wants to be 99% sure that they have enough cheese on any given day, how much cheese should they have on hand each day?

35. A type of steel shaft is produced by a process that results in a mean diameter of 4.10 cm with a standard deviation of 0.05 cm. The shaft diameters are approximately normally distributed.

    (a) What is the probability that a randomly selected shaft has a diameter in excess of 4.23 cm?
    (b) What is the probability that a randomly selected shaft has a diameter less than 4.00 cm?
    (c) What is the 10th percentile of shaft diameters?
    (d) What is the 90th percentile of shaft diameters?
    (e) Suppose a shaft is usable if it has a diameter between 4.00 and 4.20 cm, otherwise it must be scrapped. What proportion of shafts created by this process must be scrapped?

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. Which of the following statements are true?

   (a) We view statistics as random variables that have probability distributions.
   (b) We view parameters as random variables that have probability distributions.
   (c) The sampling distribution of a statistic is the probability distribution of the statistic.
   (d) In repeated sampling, the value of a statistic will vary about the parameter it estimates.
   (e) In repeated sampling, the value of a parameter will vary about the statistic that estimates it.

2. Suppose we are about to sample 2 values (without replacement) from a population that contains only 4 values. Unknown to us, the entire population is:

$$12, 13, 15, 16$$

   (a) What is the mean of the population?
   (b) How many possible samples of size $n = 2$ are there?
   (c) List all the possible samples of size 2, their probabilities of occurring, and the value of the sample mean for each sample.
   (d) What is the sampling distribution of the sample mean in this situation?
   (e) What is the mean of the sampling distribution of the sample mean?

3. Consider again the information in Question 2.

   (a) What is the value of the sample variance for each possible sample?
   (b) What is the sampling distribution of the sample variance in this situation?
   (c) If we draw a sample of size $n = 2$ from the population given in Question 2, what is the probability that the sample variance will equal 0.5?

4. Let X be the the number of heads that come up when a fair coin is tossed once, and let $\bar{X}$ be the mean number of heads per toss when a fair coin is tossed twice.

   (a) What is the probability distribution of X?
   (b) What is the mean of the probability distribution of X?
   (c) What is the variance of the probability distribution of X?
   (d) What is the probability distribution of $\bar{X}$?
   (e) What is the mean of the probability distribution of $\bar{X}$?
   (f) What is the variance of the probability distribution of $\bar{X}$?

5. Tim Horton's coffee shop chain has long held a "Roll up the Rim" contest, where purchasers of a hot beverage can roll up the rim of the cup, possibly revealing that the cup is a prize winner. The probability of winning a prize on each cup changes slightly from year to year, but it is often approximately $\frac{1}{6}$.

   (a) If you purchase 2 cups, what is the distribution of the number of winning cups?[1]
   (b) If you purchase 2 cups, what is the sampling distribution of the sample proportion of winning cups? (The sample proportion is often represented by $\hat{p}$.)

7.2 The Sampling Distribution of the Sample Mean

6. Suppose that 16 observations are randomly selected from a normally distributed population where $\mu = 10$ and $\sigma^2 = 625$.

   (a) What is the mean of the sampling distribution of $\bar{X}$?
   (b) What is the standard deviation of the sampling distribution of $\bar{X}$?
   (c) What is the distribution of $\bar{X}$?
   (d) What is $P(\bar{X} > 30)$
   (e) What is $P(0 < \bar{X} < 20)$

[1] While the probability of winning changes a tiny amount from cup to cup, as cups are revealed as winners or non-winners, there are millions of cups so the changes in probability are minuscule. For the purposes of these questions, assume that the cups can be considered independent.

7. Suppose that we intend to sample 400 observations from a large, normally distributed population. Unknown to us, the population mean and standard deviation are both equal to 20. One of the plots in Figure 7.1 is the sampling distribution of the sample mean in this scenario. Which one?

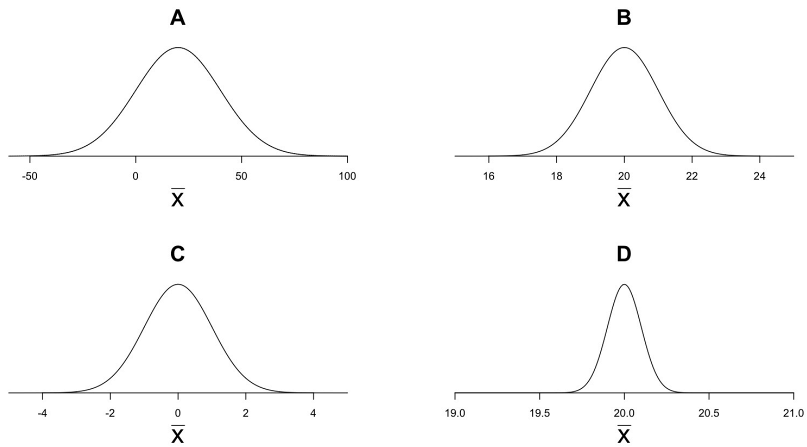


Figure 7.1: Which plot represents the sampling distribution of the sample mean?

7.3 The Central Limit Theorem

8. What does the central limit theorem tells us about the sampling distribution of the sample mean? Why is this important in the world of statistics?

9. Consider the statement: The sampling distribution of μ is normal, provided n is large. Is this statement true or false? Briefly justify your response.

10. We know that the sampling distribution of $\bar{X}$ has a mean of μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$. Which of the following statements are true?

    (a) The standard deviation of the sampling distribution of the sample mean depends on the value of μ.
    (b) We cannot possibly determine any characteristics of a statistic's sampling distribution without repeatedly sampling from the population.
    (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.
    (d) If the sample size is quadrupled, then the standard deviation of the sampling distribution of the sample mean decreases by a factor of 2.

7.4 Some Terminology Regarding Sampling Distributions

7.4.1 Standard Errors

11. What is meant by the term *standard error of a statistic*? Does the term *standard error of a parameter* have a similar meaning?

12. Suppose we randomly sample 6 values from a normally distributed population and find that $\bar{x} = 662.2$ and $s = 13.7$. What is the standard error of the sample mean?

13. All else being equal, what is the effect of the following on $SE(\bar{X})$?

    (a) An increase in the sample size.
    (b) An increase in the sample variance.
    (c) An increase in the sample mean.

7.4.2 Unbiased Estimators

14. What is meant by the term *unbiased estimator*?

15. Is the sample mean an unbiased estimator of the population mean?

7.5 Chapter Exercises

7.5.1 Basic Calculations

7.5.2 Concepts

16. In this chapter we have learned that when we randomly sample n observations from a large population that has a mean of μ and a standard deviation of σ, the sampling distribution of $\bar{X}$ has a mean of μ and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

    Which of the following statements are true?

    (a) The letter n represents the number of samples that are drawn in repeated sampling, and not the number of observations in each sample.
    (b) In practice, we usually take a large number of samples so that we can accurately estimate the sampling distribution of $\bar{X}$.

(c) The standard deviation of the population from which we are sampling decreases as the sample size increases.
(d) The standard deviation of the sampling distribution of $\bar{X}$ decreases as the sample size increases.

17. Let X represent the number on the top face when an ordinary six-sided die is rolled once, and let $\bar{X}$ represent the average of the two numbers that come up when an ordinary six-sided die is rolled twice.

    (a) What is the probability distribution of X?
    (b) What is $E(X)$?
    (c) What is σ^2?
    (d) What is the probability distribution of $\bar{X}$?
    (e) What is the mean of the sampling distribution of $\bar{X}$? (There is a quick way to find this and a slower way. Find the value using both methods.)
    (f) What is the standard deviation of the sampling distribution of $\bar{X}$? (There is a quick way to find this and a slower way. Find the value using both methods.)

18. In Question 17, X represented the number that came up when an ordinary six-sided die is rolled once, and it was found that $E(X) = 3.5$ and $Var(X) = \frac{35}{12}$. Let $\bar{X}$ represent the mean of the 2000 numbers that come up when the die is rolled 2000 times.

    (a) What is the mean of $\bar{X}$?
    (b) What is the standard deviation of $\bar{X}$?
    (c) What does the central limit theorem tell us about the distribution of $\bar{X}$?
    (d) Use the central limit theorem to approximate $P(\bar{X} < 3.5190)$.

19. 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) If we are sampling from a normally distributed population, then the sampling distribution of $\bar{X}$ is normal for any sample size.
    (b) The standard deviation of the sampling distribution of the sample mean increases 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.
    (d) If the sampling distribution of $\bar{X}$ is approximately normal, then the population from which we are sampling must be approximately normal.
    (e) All else being equal, the mean of the sampling distribution of $\bar{X}$ decreases as the sample size increases.

20. 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) 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.
(b) In practice, we usually know the true standard deviation of the sampling distribution of $\bar{X}$.
(c) In practice, we usually know the true value of μ.
(d) The sample mean is an unbiased estimator of the population mean.
(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.

21. 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) Statistics have sampling distributions.
    (b) The value of a parameter does not vary from sample to sample.
    (c) The value of a statistic does not vary from sample to sample.
    (d) All else being equal, the standard deviation of the sampling distribution of the sample mean will be smaller for $n = 10$ than for $n = 40$.
    (e) The sampling distribution of μ is usually approximately normal for $n > 30$.

7.5.3 Applications

22. A study[2] found that the amount of riboflavin (vitamin B2) in large eggs is approximately normally distributed with a mean of 0.25 mg and a standard deviation of 0.02 mg.

    (a) What proportion of large eggs have less than 0.20 mg of riboflavin?
    (b) What is the distribution of the mean amount of riboflavin in 30 large eggs?
    (c) What is the probability that the average amount of riboflavin in 30 large eggs lies between 0.240 and 0.260 mg?
    (d) What is the probability that the average amount of riboflavin in 30 large eggs lies between 0.250 and 0.255 mg?
    (e) What is the 20th percentile of the distribution of the average amount of riboflavin in 30 large eggs?

23. The amount of protein in a popular breakfast sandwich from a large fast chain is approximately normally distributed with a mean of 17.0 grams and a standard deviation of 0.80 grams. (See footnote 2.)

[2] The mean and standard deviation in this question are estimates based on information from the USDA National Nutrient Database. Assume that they are the true values for the purposes of this question.

(a) If one of these breakfast sandwiches is randomly selected, what is the probability it contains less than 16.0 grams of protein?
(b) What is the distribution of the total amount of protein in 2 randomly selected breakfast sandwiches of this type?
(c) What is the distribution of the mean amount of protein in 2 randomly selected breakfast sandwiches of this type?
(d) If 2 breakfast sandwiches of this type are randomly selected, what is the probability the total amount of protein is less than 32.0 grams?
(e) If 2 breakfast sandwiches of this type are randomly selected, what is the probability the mean amount of protein is at least 16.5 grams?
(f) What is the 97.5th percentile of the distribution of the average amount of protein in 2 randomly selected sandwiches of this type?

24. A study[3] found that the length of the right ear in Caucasian Italian men between the ages of 18 and 30 is approximately normally distributed with a mean of 62 mm and a standard deviation of 4 mm.

(a) If one Caucasian Italian man in this age group is randomly selected, what is the probability the length of his right ear is at least 60.0 mm?
(b) What is the 90th percentile of the length of the right ear in Caucasian Italian men of this age group?
(c) If 100 Caucasian Italian men are randomly selected, what is the distribution of the average length of their right ears?
(d) If 100 Caucasian Italian men are randomly selected, what is the probability that the average length of their right ears is greater than 61.8 mm?
(e) What is the 10th percentile of the distribution of the average right ear length in 100 randomly selected Caucasian Italian men?

25. A study[4] found that that carapace lengths of adult male crayfish (*Astacus leptodactylus*) in a lake in Turkey are approximately normally distributed with a mean of 48.0 mm and a standard deviation of 2.5 mm.

(a) If a single adult male crayfish from this lake is randomly selected, what is the probability its carapace length is less than 50.0 mm?
(b) What is the distribution of the mean carapace length of 10 randomly selected adult male crayfish from this lake?
(c) If 10 adult male crayfish are randomly selected from this lake, what is the probability their average carapace length is less than 50.0 mm?
(d) If 10 adult male crayfish are randomly selected from this lake, what is the

[3] Sforza et al. (2009). The mean and standard deviation given in this question are estimates based on this article, but for the purposes of this question assume that $\mu = 62$ and $\sigma = 4$.
[4] **?**. The mean and standard deviation given in this question are estimates based on this article, but for the purposes of this question assume that $\mu = 48.0$ and $\sigma = 2.5$.

probability their average carapace length lies between 46.0 mm and 50.0 mm?
(e) What is the 95th percentile of the distribution of the average carapace length of 10 randomly selected adult male crayfish from this lake?

26. A study[5] found that birth weights of female African elephants born in captivity are approximately normally distributed with a mean of 95.1 kg and a standard deviation of 13.7 kg. (The following questions are all in reference to African elephants born in captivity. The distribution of weights of elephants born in the wild is likely different from that of those born in captivity.)

    (a) What is the probability that a randomly selected newborn female African elephant weighs more than 100.0 kg?
    (b) What is the sampling distribution of the mean weight of 3 randomly selected newborn African elephants?
    (c) What is the probability that the mean weight of 3 randomly selected newborn female African elephants is more than 100.0 kg?
    (d) What is the probability that the mean weight of 10 randomly selected newborn female African elephants lies between 98.0 and 102.0 kg?
    (e) What is the 30th percentile of birth weight of female African elephants?
    (f) What is the 30th percentile of the mean birth weight of 10 randomly selected newborn female African elephants?

27. Many studies have investigated the weight of a human organs, often basing the findings on post-mortem results from accidental deaths. In one study[6], it was found that the human heart in adult Caucasian males had a mean of 365.0 grams and a standard deviation of 71.0 grams. For the purposes of this question, assume that the heart weight of adult Caucasian males is approximately normally distributed with a mean of 365.0 grams and a standard deviation of 71.0 grams.

    (a) What is the value of the 90th percentile of heart weight for adult Caucasian males?
    (b) Suppose 9 adult Caucasian males are randomly selected. What is the sampling distribution of their mean heart weight?
    (c) If 9 adult Caucasian males are randomly selected, what is the probability the average weight of their hearts is greater than 350.0 grams?
    (d) If 9 adult Caucasian males are randomly selected, what is the probability the average weight of their hearts is between 340 and 380.0 grams?
    (e) If 9 adult Caucasian males are randomly selected, what is the 90th percentile of the sampling distribution of their mean heart weight?

[5] Dale (2010). The mean and standard deviation given in this question are estimates based on this article, but for the purposes of this question assume that $\mu = 95.1$ and $\sigma = 13.7$.

[6] Lorin de la Grandmaison et al. (2001). Organ weight in 684 adult autopsies: new tables for a Caucasoid population. *Forensic Science International*, 119:149–154.

(f) If 4 adult Caucasian males are randomly selected, what is the probability that the sum of the weights of their hearts is greater than 1400 grams?

28. A study found that heights of Canadian females aged 20 to 39 are approximately normally distributed with a mean of 163.3 cm and a standard deviation of 6.4 cm.[7]

    The heights of adult females in Canada are approximately normally distributed with a mean of 164.0 cm, and a standard deviation of 6.9 cm.

    (a) What is the probability that a randomly selected Canadian female in this age group is at least 170.0 cm tall?
    (b) If 100 Canadian females in this age group are randomly selected, what is the sampling distribution of their average height?
    (c) If 100 Canadian females in this age group are randomly selected, what is the probability their average height is less than 165.0 cm?
    (d) If 100 Canadian females in this age group are randomly selected, what is the probability their average height is between 163.0 and 165.0 cm?
    (e) For a random sample of 100 Canadian females in this age group, what are the 2.5th and 97.5th percentiles of the sampling distribution of their average height?

7.5.4 Extra Practice Questions

29. Suppose that marks on a test in a very large statistics course are approximately normally distributed with a mean of 71 and a variance of 100. Suppose that we are about to randomly sample 5 of the students that wrote the test.

    (a) What is the mean of the sampling distribution of the mean score of the 5 students?
    (b) What is the standard deviation of the sampling distribution of the mean score of the 5 students?
    (c) What is the distribution of the sampling distribution of the mean score of the 5 students?
    (d) If a single student is randomly selected, what is the probability they scored higher than 75.0?
    (e) If 5 students are randomly selected, what is the probability their mean score is greater than 75.0?
    (f) If 5 students are randomly selected, what is the probability they all scored greater than 75.0?

[7] The mean and standard deviation are estimates based on information from the Canadian Health Measures Survey (Statistics Canada, 2013). Assume these are the true values for the purposes of this question.

30. A type of steel shaft is produced by a process that results in the shaft diameters being approximately normally distributed with a mean of 4.10 cm and a standard deviation of 0.05 cm.

    (a) What is the probability that a randomly selected shaft has a diameter less than 4.23 cm?
    (b) If 16 shafts are randomly selected, what is the sampling distribution of their mean diameter?
    (c) If 100 shafts are randomly selected, what is the sampling distribution of their mean diameter?
    (d) If 5 shafts are randomly selected, what is the probability their mean diameter lies between 4.05 and 4.15 cm?
    (e) If 5 shafts are randomly selected, what is the probability their mean diameter is at least 4.15 cm?
    (f) What is the 90th percentile of the distribution of the mean diameter of 20 randomly selected shafts?

31. A supplier of pure oxygen claims that their containers of pure oxygen have a mean purity of 99.500%, with a standard deviation of 0.048%. In your first dealing with this company, your first container has a purity level of 99.43%. Assume the purity levels in these containers are approximately normally distributed.

    (a) Assuming the company's claim is true, what is the probability that a randomly selected container has a purity level less than 99.43%?
    (b) Assuming the company's claim is true, what is the probability that the mean of 10 randomly selected containers is less than 99.45%?
    (c) Assuming the company's claim is true, what is the probability that the mean of 10 randomly selected containers lies between 99.45% and 99.55%?
    (d) Suppose in your first 25 containers purchased from this company, you find a mean oxygen purity of 99.00%. Does this give strong evidence against the company's claimed purity level?

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. Suppose we are sampling from a normally distributed population and we wish to calculate a confidence interval for the population mean μ. Find the appropriate z value and margin of error in the following situations.

   (a) $n = 22$, $\sigma = 5$, 95% confidence level.
   (b) $n = 10$, $\sigma = 10$, 90% confidence level.
   (c) $n = 18$, $\sigma = 15$, 99% confidence level.
   (d) $n = 18$, $\sigma = 15$, 63.2% confidence level.
   (e) $n = 50$, $\sigma = 35$, 21.2% confidence level.

2. Suppose we draw a random sample of n observations from a normally distributed population, and we wish to construct a confidence interval for the population mean μ. Suppose that the population standard deviation σ is known. Give the confidence level for each of the following intervals.

   (a) $\bar{X} \pm 1.21\frac{\sigma}{\sqrt{n}}$
   (b) $\bar{X} \pm 1.35\frac{\sigma}{\sqrt{n}}$
   (c) $\bar{X} \pm 1.57\frac{\sigma}{\sqrt{n}}$

(d) $\bar{X} \pm 2.45\frac{\sigma}{\sqrt{n}}$
(e) $\bar{X} \pm 3.15\frac{\sigma}{\sqrt{n}}$

3. 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 statistical inference, we often construct confidence intervals for statistics.
   (b) A confidence interval is a range of plausible values for a parameter.
   (c) The confidence level of an interval is determined by sample data.
   (d) The point estimate of μ lies at the midpoint of the confidence interval for μ.
   (e) The confidence interval procedures of this section assume a normally distributed population, but this assumption becomes less important as the sample size increases.

8.2.1 Interpretation of the Interval

4. Researchers are interested in estimating the population mean body mass index (BMI) for female students at a university. (A person's BMI is their weight in kilograms divided by the square of their height in metres.) Measurements on 200 female student volunteers at this university resulted in an average BMI of 25.7 kg/m^2. (For the purposes of this question, assume that the population standard deviation is 6.0 kg/m^2. (This value is likely not far from reality.) Also assume that the volunteers can be thought of as a random sample of female students from this university. This assumption may or may not be reasonable, depending on how the volunteers were obtained.)

   (a) What is a 95% confidence interval for the population mean?
   (b) What is an appropriate interpretation of this interval?
   (c) What is a 99% confidence interval for the population mean?
   (d) What is an appropriate interpretation of this interval?

5. A study by Sforza et al. (2009) investigated physical characteristics of the ears of Caucasian Italians. In one part of the study, a sample of 66 females between the ages of 18 and 30 years old revealed a mean right ear length of 61.9 mm with a standard deviation of 4.1 mm. The resulting 95% confidence interval for μ was found to be (60.91, 62.89). Assuming the 66 females in the sample can be thought of as a random sample of Caucasian Italian females between the ages of 18 and 30, give an appropriate interpretation of this confidence interval in the context of the situation at hand.

6. Suppose that at a certain university, a random sample of 50 graduating students yielded an average student loan debt of $28,500, with an associated 95% confidence

interval for μ of (26,500, 30,500). Which, if any, of the following options are correct interpretations of this interval?

(a) 95% of graduating students at this university have student loan debt between \$26,500 and \$30,500.
(b) We can be 95% confident that the average student loan debt of the 50 students in the sample lies between \$26,500 and \$30,500.
(c) We can be 95% confident that the true mean student loan debt of graduating students at this university lies between \$26,500 and \$30,500.
(d) In repeated sampling, 95% of the 95% confidence intervals would contain \$28,500.

8.2.2 What Factors Affect the Margin of Error?

7. All else being equal, what is the effect of the following on the margin of error of a confidence interval for μ?

   (a) An increase in the sample size.
   (b) An increase in the variance.
   (c) An increase in the sample mean.
   (d) An increase in the confidence level.

8.2.3 An Example

8.3 Confidence Intervals for μ When σ is Unknown

8.3.1 Introduction

8. Suppose we are sampling from a normally distributed population and we wish to calculate a confidence interval for the population mean μ. Find the appropriate t value, standard error of $\bar{X}$, and margin of error in the following scenarios.

   (a) $n = 14$, $s = 10$, and an 80% confidence level.
   (b) $n = 22$, $s = 10$ and an 80% confidence level.
   (c) $n = 8$, $s = 2,500$ and a 90% confidence level.
   (d) $n = 8$, $s = 2,500$ and a 95% confidence level.
   (e) $n = 8$, $s = 3.2$, and a 99% confidence level.

9. 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 variance of the t distribution is greater than the variance of the standard normal distribution.
(b) The mean of the t distribution is equal to the mean of the standard normal distribution.
(c) The t distribution has more area in the tails and a lower peak than the standard normal distribution.
(d) As the degrees of freedom increase, the t distribution tends toward the standard normal distribution.
(e) The t distribution is equivalent to the standard normal distribution if the degrees of freedom are at least 30.

10. Suppose that we should be using $t_{\alpha/2}$ in the formula to find a confidence interval for μ, but we mistakenly use $z_{\alpha/2}$ in its place. Would the interval found using $z_{\alpha/2}$ be wider or narrower than if we had used the appropriate $t_{\alpha/2}$ value?

8.3.2 Examples

11. As part of an investigation by Fink et al. (2007), 32 male student volunteers at a university in Germany had their hand-grip strength measured (in kilograms force (kgf)). The results are illustrated in Figure 8.1. The sample mean and standard deviation were found to be 50.91 kgf and 7.74 kgf, respectively.

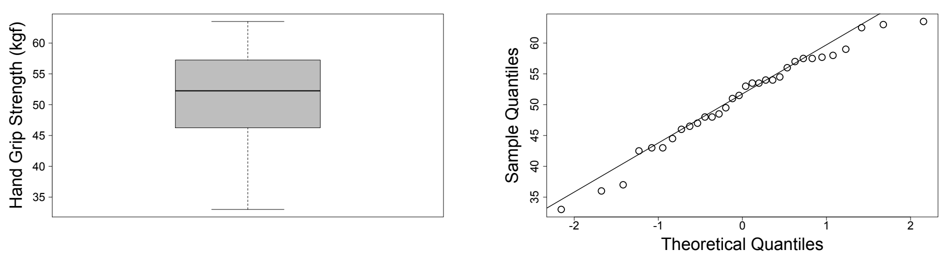


Figure 8.1: Plots of the grip strengths of 32 male student volunteers.

For the purposes of this question, assume that the volunteers can be thought of as a random sample of male students at this German university. (This may or may not be a reasonable assumption, depending on the specifics of how the volunteers were recruited.)

(a) Do these plots show a violation of the normality assumption?
(b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?
(c) What is a 95% confidence interval for the population mean μ?

(d) Give an appropriate interpretation of the 95% confidence interval in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

8.3.3 Assumptions of the One-Sample t Procedures

12. Suppose we intend to use the t procedure to construct a confidence interval for the population mean μ. If the population is not normally distributed, what are the implications?

13. What types of violation of the normality assumption are very problematic for the t procedure? What types of violation of the normality assumption are not a big problem?

8.4 Determining the Minimum Sample Size n

14. Suppose we are about to sample from a normally distributed population, and we wish to estimate the population mean μ. Calculate the minimum sample size required in the following scenarios.

    (a) $\sigma = 10$, and we wish to estimate μ to within 5 with 90% confidence.
    (b) $\sigma = 10$, and we wish to estimate μ to within 5 with 95% confidence.
    (c) $\sigma = 30$, and we wish to estimate μ to within 5 with 95% confidence.
    (d) $\sigma = 10$, and we wish to estimate μ to within 0.1 with 99% confidence.
    (e) $\sigma^2 = 16$, and we wish to estimate μ to within 1 with 95% confidence.

8.5 Chapter Exercises

8.5.1 Basic Calculations

15. Suppose we are about to randomly sample 16 values from a population where $\mu = 5$ and $\sigma = 10$. What is the value of $\sigma_{\bar{X}}$?

16. A random sample of 16 independent observations had a mean of 20.0 and a standard deviation of 2.0. What is the value of $SE(\bar{X})$?

17. Suppose the following sets of values represent samples drawn from normally distributed populations. For each scenario, calculate a 95% confidence interval for the population mean μ.

(a) 3, 5, 12.
(b) −4, 8, 23.
(c) 2012, 2010, 2054, 2032.
(d) −52, −48, −48, −46, −42.

8.5.2 Concepts

18. A person is interested in estimating the average weight (in kg) of adult males of a certain breed of dog. They draw a sample of adult males of this species, measure their weights, and find that the 95% confidence interval for the true mean weight is $(12.2, 14.9)$. Of the following options, which one is the best interpretation of this interval? (Assume that the sample can be thought of as a simple random sample of adult males of this breed.)

    (a) In repeated sampling, 95% of the time μ will lie between 12.2 kg and 14.9 kg.
    (b) In repeated sampling, 95% of the sample mean weights will lie between 12.2 kg and 14.9 kg.
    (c) In repeated sampling, 95% of the 95% confidence intervals will be $(12.2, 14.9)$.
    (d) We can be 95% confident that the population mean weight of all adult males of this breed lies between 12.2 kg and 14.9 kg.
    (e) 95% of adult males of this breed weigh between 12.2 kg and 14.9 kg.

19. Suppose we repeatedly draw random and independent samples from a normally distributed population and for each sample we calculate a 95% confidence interval for the population mean μ. In 100 of such intervals:

    (a) What is the distribution of the number of intervals that contain μ?
    (b) What is the probability that at least one interval does not contain μ?

20. Suppose that a researcher is interested in estimating the average survival time of rats after they have been infected with a certain type of virus. The researcher infects 100 rats with the virus and then measures the time until death. The average time until death is found to be 1.4 months. (For the purposes of this question, assume that σ is known to be 1.0 months.)

    (a) Survival times are often skewed to the right. If that is the case here, would inference methods based on the normal distribution be inappropriate?
    (b) What is the the value of the standard deviation of the sampling distribution of the sample mean?
    (c) What is a 92% confidence interval for the mean μ?
    (d) What is the true value of μ?
    (e) What does μ represent in this situation?

21. True or false? $t_{\alpha/2}$ tends toward $z_{\alpha/2}$ as the degrees of freedom increase.

22. (Warning: This is a very challenging question!)
In a (not very helpful) video on YouTube, a statistics instructor at a university talks about interpreting a confidence interval for the population mean, and he states that if we were to repeatedly sample from the population, 95% of the *sample* means would fall within the 95% confidence interval. This is simply untrue. A correct interpretation is that in repeated sampling, 95% of the calculated intervals would capture the true value of the population mean μ.

    This question was motivated by that YouTube video. Suppose that we are about to draw a simple random sample of n observations from a normally distributed population, then calculate a 95% confidence interval for μ. (Suppose that σ is known, so we will calculate the interval using the formula: $\bar{X} \pm z_{.025} \frac{\sigma}{\sqrt{n}}$.) If we were to draw another, independent sample of the same size from this population, what is the probability that the sample mean of this second sample falls within the 95% confidence interval for μ found in the first sample? (The probability is not 0.95, as you may guess from the fact that this question is being asked.)

23. (Warning: This question is not very complicated, but most students find it conceptually difficult. If you don't understand it, don't spend too much time on it.)
A researcher wants to calculate a 95% confidence interval for a parameter θ. The statistic $\hat{\theta}$ is a point estimator of θ, and is more complicated than any estimator that we have discussed. The researcher draws a random sample ($n = 300$), and finds that $\hat{\theta} = 2.00$ for her sample. Suppose it is known that for samples of this size, the sampling distribution of $\hat{\theta}$ is normal with a mean of θ and a standard deviation of 3.00. What is a 95% confidence interval for θ?

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.
    (b) The one-sample t procedures are robust to violations of the normality assumption.
    (c) The t procedures do not perform well when there is strong skewness or outliers in the data, especially for small sample sizes.
    (d) The width of a confidence interval for μ decreases as the sample size decreases.
    (e) The width of a confidence interval for μ decreases 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.

(b) Standard errors are the estimated standard deviations of parameters 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%.
(d) All else being equal, the width of a confidence interval decreases as the confidence level increases.
(e) All else being equal, the margin of error decreases 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.
(g) All else being equal, the confidence level of an interval increases as the sample size increases.

8.5.3 Applications

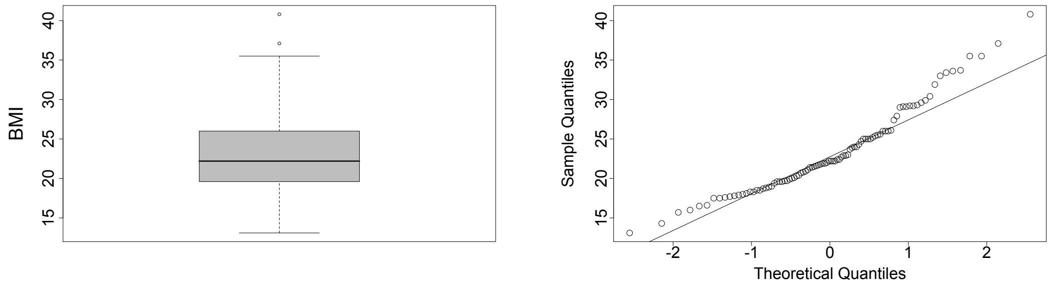
26. Body mass index (BMI $= \frac{\text{Weight}}{\text{Height}^2}$, where weight is in kg and height is in m) is a measure of body shape. A study[1] investigated BMI in Texas youth. Student participants were recruited from 6 schools in a school district in a rural area of North Central Texas (about 100 km from Fort Worth). Students participated only if their parents provided informed consent. Students who participated were given a free t-shirt.

Here will look only at the BMI data for the 94 14 year-old boys in the study. Figure 8.2 illustrates the BMI values for the 94 boys. For the boys in this sample, the mean BMI is 23.3, with a standard deviation of 5.4.

(a) Do these plots show a violation of the normality assumption? If there is a violation, is it a serious problem in this situation? (Can we still use the t procedures?)
(b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?
(c) What is a 95% confidence interval for the population mean μ?
(d) Give an appropriate interpretation of the 95% confidence interval in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

27. Franklin et al. (2012)[2] investigated various aspects of tandem running in ants. Tan-

[1] Duran et al. (2013). Growth and weight status of rural texas school youth. *American Journal of Human Biology*, 25:71–77. The data used here is simulated data with the same summary statistics as given in their Table 1.
[2] Franklin et al. (2012). Do ants need to be old and experienced to teach? *The Journal of Experimental*

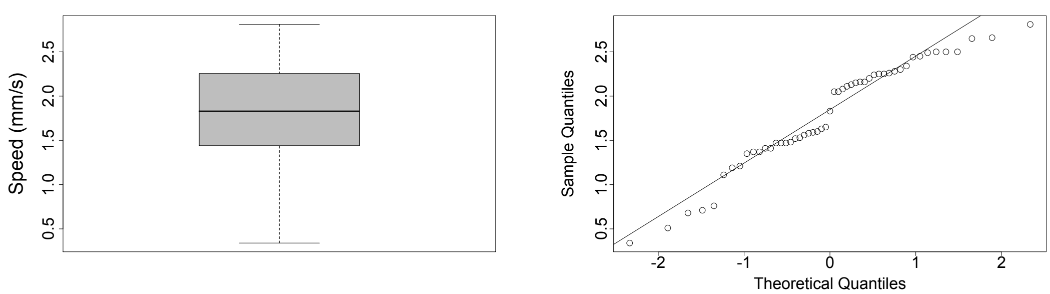


(a) Boxplot. (b) Normal quantile-quantile plot.

Figure 8.2: BMI for 94 boys in rural Texas.

dem running is a form of recruitment in which one ant with knowledge of the location of a food source or new nest site leads another ant to that location. (Optional: Watch an example of tandem running here: (1:53) (http://www.youtube.com/watch?v=X2C7Sy2oPik)) Tandem running can be thought of as a form of teaching and learning.

The study investigated various factors associated with tandem running, and investigated whether the age of the ant and the experience of the ant had any effect on tandem running. Ants (*Temnothorax albipennis*) were categorized into 4 categories: Young and inexperienced (YI), young and experienced (YE), old and inexperienced (OI), old and experienced (OE). One aspect of the study investigated the speed of tandem running for 51 tandem runs in which a YE ant was leading a YI ant. The mean speed (mm/s) of the run was recorded. Figure 8.3 illustrates the speed of the runs. For these 51 runs, the sample mean was 1.65 mm/s and the sample standard deviation was 0.62 mm/s.



(a) Boxplot. (b) Normal quantile-quantile plot.

Figure 8.3: Speed of 51 tandem runs.

(a) Do these plots show a violation of the normality assumption? If there is a

Biology, 215:1287–1292 The data used here is simulated data based on their Figure 2, with similar results and conclusions.

violation, is it a serious problem in this situation? (Can we still use the t procedures?)

(b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?

(c) What is a 95% confidence interval for the population mean μ?

(d) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

28. The *Athens Collection* is a collection of 225 skeletal remains, housed at the University of Athens in Greece. The remains were collected from cemeteries near Athens, from people who lived mainly in the second half of the 20th century. A study[3] investigated the diameter of molar teeth of individuals in this collection. Of 24 upper M1 molars from adult males, the mean crown diameter was found to be 10.38 mm with a standard deviation of 0.63 mm.

    (a) Suppose we wish to construct a 95% confidence interval for the mean crown diameter of upper M1 molars in adult males in Greece. What assumptions are necessary for the t confidence interval procedure to be valid? Are the assumptions likely to be satisfied in this scenario?
    (b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?
    (c) Assuming normality, construct a 95% confidence interval for the population mean.
    (d) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

29. Larviposition is the laying of larvae (the eggs hatch inside the female), and it occurs in a small number of insect species. Cook et al. (2012) investigated various aspects of larviposition of the blowfly *Calliphora varifrons*. In one aspect of the study, female fecundity was investigated by trapping female *C. varifrons* blowflies in a field, dissecting them, and counting the number of live larvae they had inside. Of the 49 females that carried live larvae, the mean number of larvae was 33.4, with a standard deviation of 7.0.

    (a) Suppose we wish to construct a 95% confidence interval for the mean number of live larvae carried by female *C. varifrons* blowflies in this field. What assumptions are necessary for the t confidence interval procedure to be valid? Are the assumptions likely to be satisfied in this scenario?
    (b) Although the full data was not reported in the original journal article, the

[3] Zorba et al. (2013). Evaluation of the accuracy of different molar teeth measurements in assessing sex. *Forensic Science, Medicine, and Pathology*, 9:13–23.

authors did report that the number of live larvae varied between 20 and 44 for the 49 females in the study. What does this tell us about the validity of the t procedure in this scenario?

(c) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?

(d) Assuming normality, construct a 95% confidence interval for the population mean.

(e) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

30. In another aspect of the study first discussed in Question 29, Cook et al. (2012) presented 42 female *C. varifrons* blowflies with a cube of fresh beef liver, and measured the rate of larviposition. The rate of deposition (number of larvae laid per second) had an average of 0.46, with a standard deviation of 0.32.

    (a) Suppose we wish to construct a 95% confidence interval for the true mean rate of larvae deposition for female *C. varifrons* blowflies under the conditions of this experiment. What assumptions are necessary for the t confidence interval procedure to be valid? Are the assumptions likely to be satisfied in this scenario?

    (b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?

    (c) Assuming normality, construct a 95% confidence interval for the population mean.

    (d) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

31. The nutrition information published by a popular fast-food chain claims that, in their U.S. locations, a serving of hash browns contains 310 mg of sodium. A sample of 6 servings of these hash browns resulted in a mean sodium content of 324.8 mg and a standard deviation of 40.0 mg.[4]

    (a) Suppose we wish to construct a 95% confidence interval for the true mean sodium content in servings of hash browns from U.S. locations of this fast-food chain. What assumptions are necessary for the t confidence interval procedure to be valid? Are these assumptions likely to be satisfied in this scenario?

    (b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?

    (c) The sample size is quite small here ($n = 6$), so the normality assumption is very important. It is always a little sketchy using the t procedures for such a

[4]These values are based on sample data found in the USDA National Nutrient Database.

small sample size, but none of the observations (not shown here) were outliers, and there were no obvious problems with the data, so the use of the t procedure might be reasonable. 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, construct a 95% confidence interval for the population mean.

(d) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

(e) The chain's claimed mean sodium content is 310 mg. Based on our analysis, is this claimed value a plausible value of the population mean sodium content?

32. The ratio of the length of the index finger to the length of the ring finger is called the **2D:4D ratio**. This ratio is believed to be influenced by fetal exposure to testosterone, and it has been linked to characteristics such as aggression and to diseases such as certain types of cancer. The ratio tends to be less than 1 in both sexes (the index finger tends to be shorter than the ring finger), and the 2D:4D ratio tends to be smaller in men than in women. The distribution of the 2D:4D ratio differs between groups of different ethnicities.

    A study[5] investigated the 2D:4D ratio in samples of women drawn from a sexual health clinic in Manchester, UK. In one aspect of the study, it was found that 46 black women had a mean 2D:4D ratio (of the left hand) of 0.963 with a standard deviation of 0.034, and the distribution was found to be "close to normal".

    (a) Suppose we wish to construct a 95% confidence interval for the true mean 2D:4D ratio for black women in the Manchester area. What assumptions are necessary for the t confidence interval procedure to be valid? Are the assumptions likely to be satisfied in this scenario?

    (b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?

    (c) Construct a 95% confidence interval for the population mean.

    (d) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

33. Leth (2009) investigated various aspects of homicides in Denmark. One aspect of the study involved investigating characteristics of intimate partner homicide (a person murdered their intimate partner). In 36 intimate partner homicides over the years 1983–2007, the average age of the perpetrator was 43.8 years and the standard deviation was 15.5 years.

[5] Brabin et al. (2008). The second to fourth digit ratio (2d:4d) in women with and without human papillomavirus and cervical dysplasia. *American Journal of Human Biology*, 20:337–341

(a) Suppose we wish to construct a 95% confidence interval for the true (theoretical) mean age of the perpetrators of intimate partner homicides in Denmark over these years. What assumptions are necessary for the t confidence interval procedure to be valid? Are the assumptions likely to be satisfied in this scenario?
(b) What is the point estimate of the population mean μ? What is the value of the standard error of the sample mean?
(c) Assuming normality, construct a 95% confidence interval for the population mean.
(d) Give an appropriate interpretation of the 95% confidence interval, in the context of the situation at hand. To what population do your conclusions apply? Comment on any biases that might be present.

8.5.4 Extra Practice Questions

34. A new airline is interested in estimating the mean flying time for 747 flights from Toronto (YYZ) to Vancouver (YVR). A sample of 15 flights yielded a mean time of 295.6 minutes and a sample standard deviation of 10.2 minutes. The airline wishes to use the data to construct a 95% confidence interval for the population mean flight time. They (wisely) think that they should check the normality assumption before using a procedure based on the assumption of a normally distributed population. They plot a normal quantile-quantile plot of the 15 flight times, and the results are illustrated in Figure 8.4.

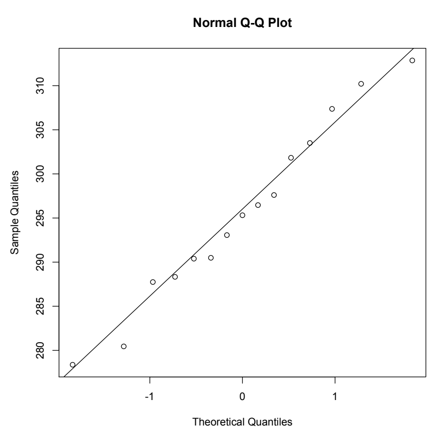


Figure 8.4: Normal QQ plot for the 15 flight times in Question 34.

(a) Does this plot give any indication that the t procedures should not be used here?

(b) Assuming that the sample of flight times can be thought of as a simple random sample from the population of flight times, and assuming that the flight times are approximately normally distributed, calculate a 95% confidence interval for the population mean flight time.

(c) Suppose that the interval in 34b was found to be (288.1, 296.1). This is not the correct interval but assume that it is for the purposes of this question. Of the following options, which one is the best interpretation of that interval?

i. 95% of YYZ–YVR flights take between 288.1 and 296.1 minutes.

ii. We can be 95% confident that the population mean flight time from YYZ to YVR for these types of flights lies between 288.1 and 296.1 minutes.

iii. We can be 95% confident that the sample mean flight time of the 15 flights lies between 288.1 and 296.1 minutes.

iv. A randomly selected YYZ–YVR flight time has a 95% chance of taking between 288.1 and 296.1 minutes.

v. All of the above.

35. Suppose you wish to estimate the average arsenic level in the soil of playgrounds of elementary schools in a very large school district. You randomly sample schools from this district, find that the sample mean arsenic level is 14.60 ppm with a 95% confidence interval for μ of $(14.1, 15.1)$. Of the following options, which one is the most appropriate interpretation of that interval?

(a) In repeated sampling, 95% of the sample means will fall between 14.1 and 15.1 ppm.

(b) We can be 95% confident that the mean arsenic level of the schools in the sample lies between 14.1 and 15.1 ppm.

(c) In repeated sampling, 95% of the time the confidence interval for μ will be (14.1, 15.1).

(d) We can be 95% confident that the population mean arsenic level of the soils of schools in this district lies between 14.1 and 15.1 ppm.

(e) In repeated sampling, μ will change from sample to sample. 95% of the time μ will lie between 14.1 and 15.1 ppm, and 5% of the time it will be outside of this interval.

36. Executives at a large company wish to investigate the average wait time for phone calls that are received at a call centre. The wait times (in minutes) for a random sample of 600 phone calls to the call centre are recorded, entered into the statistical software R, and the following output is found.

```
    One Sample t-test
data:  calltime
t = 50.0812, df = 599, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2.884044 3.119472
```

Give an appropriate interpretation of the confidence interval found in the output.

37. A car manufacturer is investigating the fuel consumption of two types of engine design. As part of the study, 8 engines of Type A are placed into a certain type of car and driven under real-world conditions. The cars are driven under similar conditions and the fuel consumption (litres/100 km) is recorded. The eight cars had a mean fuel consumption of 6.8 l/100k with a standard deviation of 0.32 l/100k.

    (a) If we had access to the data, how should we go about investigating whether the fuel consumption values are approximately normally distributed?
    (b) Suppose that it is reasonable to assume that the fuel consumption values are approximately normally distributed. What is a 95% confidence interval for the mean fuel consumption for this type of engine?
    (c) Give an appropriate interpretation of the 95% interval.

38. As part of a study into how many facial tissues should be put into a standard size box, a facial tissue company wants to investigate how many facial tissues a cold sufferer will use per day on average. They randomly sample 26 cold sufferers, watch them for a day, and find that they use an average of 29.3 facial tissues. The sample standard deviation was found to be 11.9. Assume the population is normally distributed in this situation (but keep in mind that this is a bit of a sketchy assumption).

    (a) What is the point estimate of the population mean?
    (b) What is the estimated value of the standard deviation of the sampling distribution of the sample mean? In other words, what is the *standard error* of the sample mean?
    (c) What is the appropriate 95% confidence interval for μ, the population mean number of facial tissues used per day?
    (d) What is the 90% margin of error?
    (e) What is the appropriate 90% confidence interval?
    (f) What is an appropriate interpretation of the 90% interval?

39. Food scientists developed a new type of hydrogenated vegetable oil, and they are investigating several properties of this oil. As part of the investigation, they want to know the average melting temperature of the new oil. A random sample of 10 batches of this oil are heated until they melt, and the melting temperature is then

recorded. The sample mean of the 10 observations was found to be 34.57 degrees Celsius. The output for this data is:

```
        One Sample t-test
data:  oil
t = 554.0799, df = 9, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 34.43033 34.71262
sample estimates:
mean of x
 34.57147
```

Give an appropriate interpretation of the confidence interval found in the output, *in the context of the problem at hand*.

40. Cadmium is a heavy metal that can build up to high levels in mushrooms. Researchers investigated cadmium concentration in a specific type of wild mushroom. A random sample of 10 mushrooms was obtained, and it was found that the sample mean and sample standard deviation were 3.2 and 0.6, respectively. The units are mg/kg of dry weight. Suppose it is reasonable to assume that the cadmium concentrations are normally distributed.

    (a) Calculate a 95% confidence interval for the population mean cadmium concentration for this type of mushroom.
    (b) Give an interpretation of the interval in the context of the problem.
    (c) Is it reasonable to simply assume the concentrations are normally distributed? Should this assumption be investigated? If so, how? What are the consequences if the concentrations are not normally distributed?

41. Researchers were interested in estimating the mean calcium level in the blood of women in rural Guatemala. A sample of 18 women in a certain area of rural Guatemala revealed a sample mean of 9.2 mg/dL, with a sample standard deviation of 0.38 mg/dL. Assume for the purposes of these questions that the calcium levels are approximately normally distributed, and the sample can be thought of as a simple random sample of women from the area.

    (a) What is a 95% confidence interval for the population mean?
    (b) The correct interval 95% interval is (9.01, 9.39). Which one of the following is the most appropriate interpretation of this interval?
        i. We can be 95% confident that the population mean calcium level of women in this area lies between 9.01 and 9.39 mg/dL.
        ii. We can be 95% confident that the sample mean calcium level of the 18 women in the sample lies between 9.01 and 9.39 mg/dL.
        iii. In repeated sampling, 95% of the sample means would lie between 9.01

and 9.39 mg/dL.

iv. We are more than 95% confident that the population mean lies within the interval, since 9.2 lies within the interval.

v. Since confidence intervals are two-sided, we can only be 90% confident that the population mean calcium level of women in this area lies between 9.01 and 9.39 mg/dL.

42. A researcher drew a random sample from a normally distributed population and found a sample mean of 61.1. For this particular population, the population standard deviation was known to the researcher but the population mean was unknown. The researcher found a 95% confidence interval for the population mean of (60.4, 61.8).

    (a) What is the point estimate of μ?
    (b) What is the margin of error?
    (c) Calculate a 93% confidence interval for the population mean.

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. Give appropriate hypotheses in each of the following scenarios.

   (a) You wish to investigate whether there is a difference in mean resting heart rates of male and female Olympic athletes.

   (b) You wish to investigate whether there are differences in the six-month survival rates of four different cancer treatments. (The six-month survival rate is the true proportion of patients that survive for at least six months.)

   (c) A producer of a type of protein bar claims their bars contain no more than 4.0 grams of saturated fat on average. You wish to show that the true mean amount of saturated fat exceeds 4.0 grams.

   (d) An auto manufacturer claims that at least 98% of purchasers of one of their top models are satisfied with their purchase. You wish to show that the true percentage is less than 98%.

   (e) You wish to investigate whether there are differences in the variances of the gestation periods of three species of mammal.

9.2 The Logic of Hypothesis Testing

9.3 Hypothesis Tests for μ when σ is Known

9.3.1 Constructing Appropriate Hypotheses

2. Would it make sense to test the hypothesis H_0: $\bar{X} = 2$? Why or why not?

9.3.2 The Test Statistic

3. Suppose we are interested in testing a null hypothesis about a population mean μ. This population is normally distributed with a standard deviation of $\sigma = 5$. Calculate the value of the appropriate test statistic in the following situations.

   (a) $n = 10$, $\bar{X} = 15.5$, $\mu_0 = 20$.
   (b) $n = 40$, $\bar{X} = 15.5$, $\mu_0 = 20$.
   (c) $n = 20$, $\bar{X} = 35.5$, $\mu_0 = 30$.
   (d) $n = 20$, $\bar{X} = 65.5$, $\mu_0 = 30$.

4. A researcher draws a random sample of 75 observations from a normally distributed population. The mean of the population is unknown, but the standard deviation of the population is known to be 6. The researcher wants to test the null hypothesis that the mean of the population is 12, against the alternative that it is different from 12. The researcher intends to use the test statistic:

$$\frac{\bar{X} - 12}{6/\sqrt{75}}$$

   What is the sampling distribution of this test statistic if the null hypothesis is true?

9.3.3 The Rejection Region Approach to Hypothesis Testing

5. Suppose we wish to carry out a hypothesis test of H_0: $\mu = \mu_0$ for a normally distributed population where σ is known. In the following scenarios, what values of the Z test statistic lead to a rejection of the null hypothesis? (What is the appropriate rejection region?)

   (a) H_a: $\mu \neq \mu_0$, $\alpha = 0.10$.
   (b) H_a: $\mu < \mu_0$, $\alpha = 0.10$.

(c) H_a: $\mu > \mu_0$, $\alpha = 0.10$.
(d) H_a: $\mu \neq \mu_0$, $\alpha = 0.004$.
(e) H_a: $\mu < \mu_0$, $\alpha = 0.004$.
(f) H_a: $\mu > \mu_0$, $\alpha = 0.004$.

6. 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 significance level of a test is the probability of rejecting the null hypothesis, given the null hypothesis is true.
   (b) If we reject the null hypothesis, then we know that the null hypothesis is false.
   (c) If we reject the null hypothesis, then we know that the alternative hypothesis is false.
   (d) If we do not reject the null hypothesis, then we know that the null hypothesis is true.
   (e) If we do not reject the null hypothesis, then we know that the alternative hypothesis is true.

9.3.4 P-values

7. Suppose we wish to test a null hypothesis about the mean of a normally distributed population. Find the p-value in the following scenarios.

   (a) $Z = 2.76$, H_a: $\mu \neq 10$.
   (b) $Z = 2.76$, H_a: $\mu < 10$.
   (c) $Z = 2.76$, H_a: $\mu > 10$.
   (d) $Z = -1.74$, H_a: $\mu \neq 20$.
   (e) $Z = -1.74$, H_a: $\mu < 20$.
   (f) $Z = -1.74$, H_a: $\mu > 20$.

8. Would the null hypothesis be rejected in the following scenarios?

   (a) H_a: $\mu \neq 10$, p-value $= 0.04$, $\alpha = 0.05$.
   (b) H_a: $\mu > 10$, p-value $= 0.04$, $\alpha = 0.05$.
   (c) H_a: $\mu \neq 10$, p-value $= 0.12$, $\alpha = 0.10$.
   (d) H_a: $\mu > 10$, p-value $= 0.12$, $\alpha = 0.10$.
   (e) H_a: $\mu < 10$, p-value $= 0.08$, $\alpha = 0.10$.

9. Suppose that for a certain population, the standard deviation is known but the mean is unknown. A researcher draws a random sample of 600 observations from this population, and finds a sample mean of 2200. They carry out a Z test of the null hypothesis that the population mean is 1000 against the alternative that it is

greater than 1000, and find that the p-value is 0.08. Of the following options, which one best describes the meaning of this p-value?

(a) If the null hypothesis is true, the probability of obtaining a sample mean at least as large as 2200 is 0.08.
(b) If the null hypothesis is false, the probability of obtaining a sample mean at least as large as 2200 is 0.08.
(c) The probability that the population mean is greater than 1000 is 0.08.
(d) The probability that the null hypothesis is false is 0.08.
(e) The probability that the null hypothesis is true is 0.08.

10. Test your conceptual understanding: Which of the following statements about p-values are true? (You should be able to explain *why* a statement is true or *why* a statement is false.)

(a) A p-value is a probability.
(b) We reject the null hypothesis when the p-value is less than or equal to the significance level.
(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.
(d) The p-value of a test depends on sample data and on the null and alternative hypotheses.
(e) The p-value of a two-sided test will be greater than 1 if the null hypothesis is true.

11. Test your conceptual understanding: Which of the following statements about p-values are true? (You should be able to explain *why* a statement is true or *why* a statement is false.)

(a) The p-value of a test will be less than 0.05 if the null hypothesis is false.
(b) The p-value of a test is the probability that the null hypothesis is true.
(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.
(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.
(e) If the p-value of a test is less than 0 that means there is extremely strong evidence against the null hypothesis.

9.4 Examples

12. A Slovenian study[1] investigated the concentrations of cadmium and lead in edible mushrooms in an area near the largest Slovenian thermal power plant. In one aspect of the study, a sample of caps ($n = 29$) from the mushroom *Boletus edulis* revealed a mean cadmium concentration of 9.26 mg/kg of dry weight. Suppose that for samples of this type, $\sigma = 5.00$ mg/kg dw (this is an estimate based on this study, but for the purposes of this question assume that the value of σ is known to be 5.00). Slovenian regulations give a tolerable value of 3 mg/kg dw (a cadmium level of 3 mg/kg of dry weight or less is considered acceptable for consumption). Does this sample provide strong evidence that the concentration of cadmium in this type of mushroom in this area exceeds the tolerable value? Suppose we wish to carry out a Z test of the appropriate hypothesis at the $\alpha = 0.05$ significance level. (There was some indication of a little right-skewness in the data, but the sample size here is not very small, so it's not unreasonable to use the Z test.)

    (a) Give the appropriate hypotheses in words and symbols.
    (b) What is the value of the appropriate test statistic?
    (c) What is the p-value of the test?
    (d) Is the null hypothesis rejected at the $\alpha = 0.05$ level of significance?
    (e) Give an appropriate conclusion to the hypothesis test.

13. A 2015 National Vital Statistics Report showed that in 2013, the weights of all full-term newborn babies in the United States had a mean of 3.35 kg. Does the mean weight of Hispanic newborns in the U.S. differ from this? A random sample of 100 Hispanic newborns in 2013 had a mean of 3.318 kg. Suppose the standard deviation of birth weights in the Hispanic population of the United States is known to be $\sigma = 0.47$ kg. (This standard deviation is based on sample data, but it is very close to the true value. For the purposes of this question, assume this is the correct value of the population standard deviation.) Test the null hypothesis that the mean weight of full-term Hispanic newborns in the U.S. in 2013 is the same as that of the general population (3.35 kg).

    (a) Give the appropriate hypotheses in words and symbols.
    (b) What is the value of the appropriate test statistic?
    (c) What is the p-value of the test?
    (d) Is the null hypothesis rejected at the $\alpha = 0.05$ level of significance?
    (e) Give an appropriate conclusion to the hypothesis test.

[1] Petkovšek, S. and Pokorny, B. (2013). Lead and cadmium in mushrooms from the vicinity of two large emission sources in Slovenia. *Science of the Total Environment*, 443:944–954

9.5 Interpreting the p-value

9.5.1 The Distribution of the p-value When H_0 is True

14. Suppose we are about to draw a random sample from a normally distributed population and carry out a Z test of H_0: $\mu = 10$.

    (a) If the null hypothesis is true, what is the distribution of the p-value?
    (b) If the null hypothesis is true, what is the average value of the p-value?

9.5.2 The Distribution of the p-value When H_0 is False

15. Give an interpretation of the following p-values in terms of the evidence they give against the null hypothesis. Suppose that there is no given value of α.

    (a) 2.9×10^{-48}
    (b) 0.0008
    (c) 0.02
    (d) 0.07
    (e) 0.12
    (f) 0.45
    (g) 0.88

9.6 Type I Errors, Type II Errors, and the Power of a Test

16. Suppose we are testing H_0: $\mu = 10$ against a two-sided alternative hypothesis at $\alpha = 0.05$.

    (a) Suppose $\mu = 5$, and when we carry out the test we find a p-value of 0.03. Will we make a Type I error, a Type II error, or neither error?
    (b) Suppose $\mu = 5$, and when we carry out the test we find a p-value of 0.16. Will we make a Type I error, a Type II error, or neither error?
    (c) Suppose $\mu = 10$, and when we carry out the test we find a p-value of 0.02. Will we make a Type I error, a Type II error, or neither error?
    (d) Suppose $\mu = 10$, and when we carry out the test we find a p-value of 0.28. Will we make a Type I error, a Type II error, or neither error?

17. In a hypothesis test, under what conditions should a small value of α (0.00001, say) be chosen? Under what conditions would it be more reasonable for α to be a larger value (0.20, say)?

9.6.1 Calculating Power and the Probability of a Type II Error

18. Suppose we are about to sample 100 observations from a normally distributed population where it is known that $\sigma = 20$, but μ is unknown. We intend to test H_0: $\mu = 30$ against H_a: $\mu < 30$ at $\alpha = 0.05$.

    (a) What values of the sample mean would lead to a rejection of the null hypothesis?
    (b) What is the power of the test if $\mu = 28$?
    (c) What is the power of the test if $\mu = 26$?

19. Suppose we are about to sample 400 observations from a normally distributed population where it is known that $\sigma = 160$, but μ is unknown. We intend to test H_0: $\mu = 800$ against H_a: $\mu > 800$ at $\alpha = 0.01$.

    (a) What values of the sample mean would lead to a rejection of the null hypothesis?
    (b) What is the power of the test if $\mu = 810$?
    (c) What is the power of the test if $\mu = 840$?

20. (Warning: This question involves a power calculation for a test with a two-sided alternative, and it is meaningful only if we do not wish to reach a directional conclusion in the hypothesis test (we intend to either *reject* H_0 or *not reject* H_0). Interpreting power can be tricky when the alternative hypothesis is two-sided and we want to reach a directional conclusion.)

    Suppose we are about to sample 50 observations from a normally distributed population where it is known that $\sigma = 12$, but μ is unknown. We intend to test H_0: $\mu = 15$ against H_a: $\mu \neq 15$ at $\alpha = 0.05$.

    (a) What values of the sample mean would lead to a rejection of the null hypothesis?
    (b) What is the power of the test if $\mu = 16$?
    (c) What is the power of the test if $\mu = 17$?
    (d) What is the power of the test if $\mu = 13$?

21. A researcher wants to test the null hypothesis that the mean of a certain population is 0, against the alternative hypothesis that it is greater than 0. He has taken a statistics course, but he forgets most of the material. He decides to take a random

sample of size 50, and if the sample mean is larger than 10 he will reject the null hypothesis. Suppose that the population is normally distributed and $\sigma = 100$.

(a) If $\mu = 0$ what is the probability the researcher commits a Type I error?
(b) If $\mu = 0$ what is the probability the researcher commits a Type II error?
(c) If $\mu = 5$ what is the probability the researcher commits a Type I error?
(d) What is of the power of the test if $\mu = 5$?
(e) What is of the power of the test if $\mu = 10$?

9.6.2 What Factors Affect the Power of the Test?

22. All other factors being equal, what is the effect of the following on the power of the test?

(a) An increase in the sample size.
(b) A decrease in the significance level.
(c) An increase in the distance between the true mean and the hypothesized mean.
(d) A decease in the population variance.

9.7 One-Sided Test or Two-Sided Test?

23. The mean of a certain population is thought to be 1000. We wish to draw a sample and test the null hypothesis that the mean is 1000. In two or three sentences, state the advantages and disadvantages of choosing a one-sided alternative hypothesis instead of a two-sided alternative.

24. Suppose it is commonly believed that the mean of a certain population is 40. You feel that the mean is actually much larger than 40, so you draw a sample and carry out a one-sided hypothesis test at the 5% level of significance, using the alternative hypothesis H_a: $\mu > 40$. You find that the sample mean is actually much smaller than 40, and you obtain a p-value of 0.99. You now decide to simply change your mind, abandon your earlier hypothesis and, using the same data, test the alternative hypothesis H_a: $\mu < 40$. You find and report a p-value of 0.01, and reject the null hypothesis at the 5% significance level. Is this acceptable statistical practice?

9.8 Statistical Significance and Practical Significance

25. Label the following statements as true or false. (You should be able to explain *why* a statement is true or *why* a statement is false.)

    (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.
    (b) If the sample size is extremely large, then even tiny, meaningless differences from the hypothesized value will be found statistically significant.
    (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.

9.9 The Relationship Between Hypothesis Tests and Confidence Intervals

26. The p-value for a two-sided hypothesis test of H_0: $\mu = 50$ is found to be 0.0233.

    (a) Would a 95% confidence interval for μ contain 50?
    (b) Would a 99% confidence interval for μ contain 50?

27. Suppose a 95% confidence interval for a population mean is found to be $(1.17, 8.23)$. For the same data, what can be said of the p-value of a two-sided test of the null hypothesis that the population mean is 0?

    (a) The p-value is greater than 0.05.
    (b) The p-value is less than 0.05.
    (c) The p-value is equal to 0.05.
    (d) It is impossible to determine the p-value with the given information.

9.10 Hypothesis Tests for μ When σ is Unknown

28. Suppose we wish to carry out a hypothesis test of H_0: $\mu = \mu_0$ for a normally distributed population where σ is unknown. In the following scenarios, what values of the t test statistic will lead to a rejection of the null hypothesis? (What is the appropriate rejection region?)

    (a) H_a: $\mu \neq \mu_0$, $\alpha = 0.10$, $n = 15$.
    (b) H_a: $\mu < \mu_0$, $\alpha = 0.10$, $n = 15$.
    (c) H_a: $\mu > \mu_0$, $\alpha = 0.10$, $n = 15$.

(d) H_a: $\mu \neq \mu_0$, $\alpha = 0.05$, $n = 27$.
(e) H_a: $\mu < \mu_0$, $\alpha = 0.001$, $n = 91$.
(f) H_a: $\mu > \mu_0$, $\alpha = 0.15$, $n = 348$.

29. Suppose we are testing a hypothesis about the mean of a normally distributed population. Find the p-value in each of the following situations.

(a) $t = 2.76$, $n = 5$, H_a: $\mu \neq 10$.
(b) $t = 2.76$, $n = 5$, H_a: $\mu < 10$.
(c) $t = 2.76$, $n = 5$, H_a: $\mu > 10$.
(d) $t = -1.84$, $n = 18$, H_a: $\mu \neq 20$.
(e) $t = -1.84$, $n = 18$, H_a: $\mu < 20$.
(f) $t = -1.84$, $n = 18$, H_a: $\mu > 20$.

9.10.1 Examples of Hypothesis Tests Using the t Statistic

30. The nutrition information published by a major fast-food chain claims that their bean burritos contain 11 grams of fat on average. A sample of 6 of these burritos were analyzed, and it was found that they contained an average of 11.49 g of fat with a standard deviation of 1.45 g. (This information is based on sample data published in the US National Nutrient Database.) Suppose we wish to test the null hypothesis that the published nutrition information is correct, and that these burritos contain 11.0 g of fat on average. (The use of the t procedures is always a little questionable for such a small sample size ($n = 6$), but there were no outliers in the data so the use the t procedure is not completely out of line.)

(a) In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(b) What is the value of the appropriate test statistic?
(c) What is the p-value of the test?
(d) Give an appropriate conclusion to the hypothesis test.
(e) What assumptions are necessary in order for the test to be valid? What biases might be present?

9.11 More on Assumptions

31. A sample of 20 values resulted in the normal quantile plot seen in Figure 9.1. Would it be reasonable to use the t procedures on this sample data?

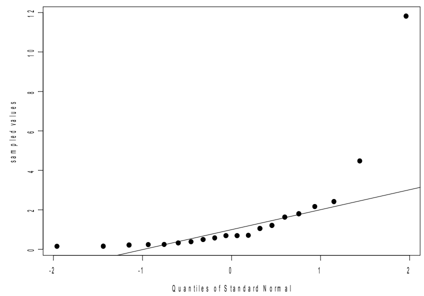


Figure 9.1: Normal QQ plot for Question 31.

32. Suppose we are about to sample from a population and carry out a two-sided t test about a population mean μ.

    (a) For what shapes of distribution does the t test perform poorly?
    (b) If we use the t test when we are sampling from a strongly skewed distribution, what are the implications?

9.12 Chapter Summary

9.12.1 Basic Calculations

33. Suppose we are interested in testing a hypothesis about the mean of a normally distributed population. Calculate the value of the appropriate test statistic in the following scenarios. Also, assuming that the null hypothesis is true, state the distribution of the test statistic in each scenario.

    (a) $n = 10$, $\bar{X} = 15.5$, $s = 2$, $\mu_0 = 20$.
    (b) $n = 10$, $\bar{X} = 15.5$, $s = 20$, $\mu_0 = 20$.
    (c) $n = 40$, $\bar{X} = 35.5$, $\sigma = 2$, $\mu_0 = 30$.
    (d) $n = 50$, $\bar{X} = 45.5$, $s = 20$, $\mu_0 = 40$.

34. Suppose we are interested in testing a hypothesis about the mean of a normally distributed population. Calculate the value of the appropriate test statistic in the following scenarios. Also, assuming that the null hypothesis is true, state the distribution of the test statistic in each scenario.

    (a) $n = 20$, $\bar{X} = 35.5$, $s = 2$, $\mu_0 = 30$.
    (b) $n = 25$, $\bar{X} = 35.5$, $\sigma = 20$, $\mu_0 = 40$.

(c) $n = 200$, $\bar{X} = 65.5$, $s = 2$, $\mu_0 = 30$.
(d) $n = 60,000$, $\bar{X} = 32.1$, $s = 10$, $\mu_0 = 32$.

35. Consider the following output for a one-sample t procedure.

```
One-sample t-Test
data:  madeupdata
t = -4.1027, df = 24, p-value = 0.0004
alternative hypothesis:  mean is not equal to 50
95 percent confidence interval:
 46.80365 48.94322
sample estimates:
 mean of x
  47.87344
```

Which of the following statements are true?

(a) The null hypothesis is $\bar{X} = 50$.
(b) The sample size is 24.
(c) The area to the left of -4.1027 under a t distribution with 24 degrees of freedom is $2 \times 0.0004 = 0.0008$.
(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.
(e) The 95% margin of error is greater than 20.0.

9.12.2 Concepts

36. What are the assumptions of a Z test on the population mean μ?

37. What are the assumptions of a t test on the population mean μ?

38. If the assumptions of a Z test or t test are violated, what are the consequences?

39. In inference procedures for a population mean μ, when is a t test more appropriate than a z test?

40. Which of the following should be based on the current sample's data?

(a) The significance level of a test.
(b) The hypothesized mean.
(c) The choice of alternative hypothesis.
(d) The sample mean.

(e) The value of the test statistic.
(f) The p-value.

41. Suppose we are about to sample from a normally distributed population and conduct a one-sample t test of H_0: $\mu = 250$.

(a) If the null hypothesis is true, what is the average value of the test statistic?
(b) If the null hypothesis is true, what is the average value of the p-value?

42. Suppose we are about to carry out 25 independent hypothesis tests. We intend to carry out each test at a significance level of $\alpha = 0.05$. Suppose that the assumptions of each test are true, and, unknown to us, the null hypothesis is true in all 25 situations.

(a) What is the distribution of the number of Type I errors that we will make in these 25 tests?
(b) On average, how many Type I errors will we make?
(c) What is the probability that we make at least one Type I error?

43. A researcher wants to test the null hypothesis H_0: $\mu = 29$ against the alternative that μ is different from 29. She forgets how to carry out the test, but remembers how to calculate confidence intervals. She finds a 95% confidence interval for μ of $(23.20, 28.80)$, and a 99% confidence interval for μ of (22.32, 29.68). Which one of the following could be the p-value of the appropriate test?

(a) 0.0022.
(b) 0.0358.
(c) 0.0517.
(d) 0.0812.
(e) 1.8143.

44. Consider the following output from a statistical computing package.

```
One-sample t-Test
data:  mydata
t = -0.2518, df = 39, p-value = 0.8025
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.3412277  0.2656748
```

(a) What is the value of the sample mean?
(b) What is the value of the margin of error?
(c) What would the p-value be if the alternative was $\mu < 0$ instead of $\mu \neq 0$? (Don't look up a range of values in a table—base your response on the output.)

(d) What would the p-value be if the alternative was $\mu > 0$ instead of $\mu \neq 0$?

(e) Suppose the p-value in the original output was 0.008025 instead of 0.8025. Give an appropriate conclusion to the t test.

45. A researcher is interested in testing the null hypothesis that the mean of a certain population is 32. She strongly believes that the population mean is less than 32, and she feels that a one-sided alternative hypothesis is appropriate. She draws a simple random sample of size 15, and uses software to perform a t-test with a one-sided alternative hypothesis. The following output summarizes the results.

```
One-sample t-Test
data:  madeupdata
t = -2.6839, df = 14, p-value = 0.0089
alternative hypothesis:  mean is less than 32
```

Consider the 3 values: 28.0, 29.0, 34.0. Which of these values could possibly be the true value of μ?

46. Consider the two normal quantile-quantile plots illustrated in Figure 9.2.

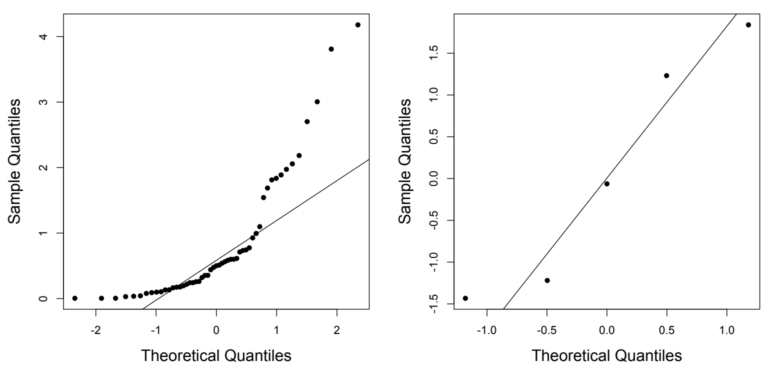


Figure 9.2

The plot on the left is for a sample of size 100. The plot on the right is for a sample of size 5. In each of these situations, would it be reasonable to use the t procedures to conduct tests and construct confidence intervals?

47. Consider the following output for a one sample problem.

```
        One Sample t-test
data:  mydata
t = 4.1611, df = 39, p-value = 0.0001687
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 51.26444 53.65645
sample estimates:
mean of x
 52.46044
```

(a) The sample standard deviation for this data is 3.739676. What is the value of the standard error of the sample mean?

(b) Give an appropriate conclusion to the hypothesis test.

(c) Suppose that before they gathered the data, the researchers in this problem decided that a 5% significance level was appropriate in this situation. Suppose also that the population is normally distributed and the null hypothesis is true. Before the data was gathered, what was the probability that this test would result in a Type I error?

(d) Suppose that the true value of the population mean is actually 50. If we make an appropriate conclusion to the hypothesis test at the 5% significance level (based on the output), would we make a Type I error, a Type II error, or neither error?

(e) The data for this problem is plotted in Figure 9.3. Do these plots give any indication that the assumptions of the t procedure are violated?

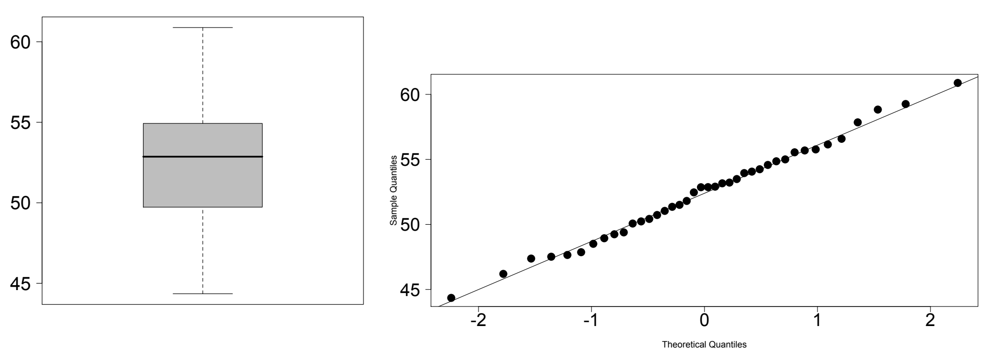


Figure 9.3: Plots for Question 47e.

48. Figure 9.4 illustrates boxplots for 3 samples (labelled A, B, and C). Each sample has a sample size of 50.

Suppose that for each sample, we carry out a hypothesis test of H_0: $\mu = 18$ against

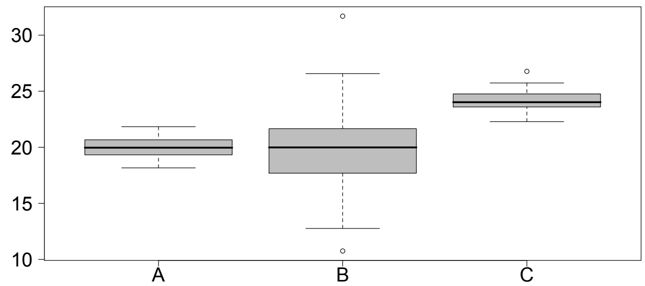


Figure 9.4: Boxplots for three samples.

a two-sided alternative hypothesis.

(a) Which one of the three samples would have the largest p-value of the test?
(b) Which one of the three samples would have the smallest p-value of the test?

49. Suppose that we run two experiments, and then carry out a hypothesis test for the data resulting from each experiment. For the data of the first experiment, the p-value is found to be 0.100000001, and for the data of the second experiment, the p-value is found to be 0.09999999999.

(a) Is the null hypothesis rejected in Experiment 1 at $\alpha = 0.10$?
(b) Is the null hypothesis rejected in Experiment 2 at $\alpha = 0.10$?
(c) Which one of the following statements is the best assessment of the strength of the evidence against the null hypothesis in the two experiments?
   i. Experiment 1 shows stronger evidence against the null hypothesis.
   ii. Experiment 2 shows stronger evidence against the null hypothesis.
   iii. The evidence against the null hypothesis is essentially the same for both experiments.

50. In hypothesis testing, under certain conditions we should choose α to be a very small value. Under which of the following situations should we choose a very small value for α?

(a) If a Type I error would cost thousands of lives, but a Type II error would result in a loss of \$2.
(b) If a Type II error would cost thousands of lives, but a Type I error would result in a loss of \$2.
(c) If we want to be very sure the null hypothesis is true.
(d) If we want the test to have very high power.

51. Suppose that you find a coin on the street that is strange looking and is obviously not regular currency. This coin may very well be a trick coin that comes up heads more (or less) often than tails. Suppose that you want to test the null hypothesis that this coin is a fair coin (the probability it comes up heads when tossed is 0.5), against the alternative that it comes up tails more often than heads on average. You foolishly decide to simply toss the coin once, and if it comes up tails, you will decide that it is biased toward tails.

    (a) If the null hypothesis is true, what is the probability of a Type I error?
    (b) Suppose that the null hypothesis is false. For what true probability of tails would it be impossible to make a Type II error?

52. The following analysis of a hypothesis test contains at least one glaring error.

    A researcher wants to test the null hypothesis H_0: $\bar{X} = 10$, against the alternative that it is less than 10. The researcher performs a t test, and observes a test statistic of -3.84. The p-value was found to be 1.12, so the null hypothesis was not rejected at the 5% level of significance.

    What errors are in this analysis?

53. A researcher draws a large sample (several hundred observations) from a normally distributed population, where it is known that $\sigma = 10$. She wants to test the null hypothesis that $\mu = 50$, against the alternative that it is less than 50. Unknown to the researcher, the true value of μ is 37. The researcher carries out a Z test, and finds that the p-value of the test is 0.5. What was the value of the sample mean? (If it is impossible to determine, say so.)

54. A pet food manufacturer suspects that one of its machines is overfilling bags of dog food that are supposed to contain 20 kg of food. They want to test the null hypothesis that the mean fill is 20 kg against the alternative hypothesis that it is greater than 20 kg. They don't know much about statistics, and decide that they will reject the null hypothesis if the sample mean of 100 randomly selected bags is greater than 20.1 kg. Unknown to the manufacturer, the machine is truly overfilling the bags, with a mean of 20.3 kg. When the manufacturer conducts the test and draws a conclusion, what is the probability that they make a Type II error? (Assume that the population standard deviation is known to be 0.5 kg.)

55. A scientist is about to carry out an experiment and then conduct a Z test on her data. She vaguely recalls how to calculate the appropriate Z test statistic, and remembers that the vast majority of randomly selected Z values fall between -3 and $+3$. She decides that she will reject the null hypothesis if the value of the Z test statistic falls outside of this range.

    (a) If the null hypothesis is true, what is the probability that she makes a Type I

error?

(b) If the null hypothesis is true (and she keeps the same rejection rule), what will happen to the probability of a Type I error as the sample size increases?
(c) If the null hypothesis is false, what is the probability that she makes a Type I error?
(d) If the null hypothesis is false (and she keeps the same rejection rule), what will happen to the power of the test as the sample size increases?

56. Suppose we should be using a t test statistic to test a null hypothesis, against a two-sided alternative. We properly calculate the t statistic, but we make a mistake find the p-value of the test from the standard normal distribution. What effect will that have on the reported p-value?

    (a) The reported p-value will be greater than it should be.
    (b) The reported p-value will be less than it should be.
    (c) The reported p-value will be equal to what it should be.
    (d) The effect on the p-value is impossible to determine.

57. 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) All else being equal, the power of a test will increase as the population variance decreases.
    (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.
    (c) The probability of a Type II error decreases the farther the true population mean is from the hypothesized mean.
    (d) If the null hypothesis is true, then we cannot make a Type II error.

58. 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) Power $= 1 - P(\text{Type I error})$.
    (b) Power $= 1 - P(\text{Type II error})$.
    (c) $P(\text{Type I error}) = 1 - P(\text{Type II error})$.
    (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.
    (e) All else being equal, the power of a test will increase as the sample size increases.

59. 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) If a null hypothesis is rejected at the 5% significance level, then it would also

be rejected at the 10% significance level.

(b) If a null hypothesis is rejected at $\alpha = 0.05$, then it would also be rejected at $\alpha = 0.01$.

(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.

(d) The significance level of a test is $1 - p$-value.

(e) The significance level of a test is the probability that the null hypothesis is true.

60. Test your conceptual understanding: Which of the following statements about p-values are true? (You should be able to explain *why* a statement is true or *why* a statement is false.)

(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.)

(b) The p-value of a test is the probability of getting an insignificant result, given the null hypothesis is true.

(c) The p-value of a test is less than the significance level whenever the null hypothesis is false.

(d) The p-value of a test is the probability of rejecting the null hypothesis when it is 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.

61. Test your conceptual understanding: Which of the following statements about p-values are true? (You should be able to explain *why* a statement is true or *why* a statement is false.)

(a) For a two-sided alternative hypothesis, the p-value of the test will equal 1 if the sample mean equals the hypothesized mean.

(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.

(c) If the null hypothesis is true, then on average the p-value will equal 0.5.

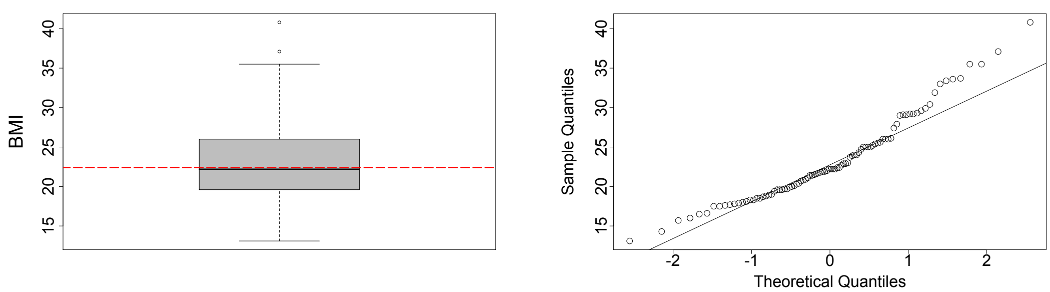
(d) If the null hypothesis is false, then on average the p-value will equal 0.5.

9.12.3 Applications

62. Body Mass Index ($\frac{\text{Weight}}{\text{Height}^2}$, where weight is in kg and height is in m) is a measure of body shape. A study[2] investigated BMI in Texas youth. Student participants were recruited from 6 schools in a school district in a rural area of North Central Texas (about 100 km from Fort Worth). Students participated only if their parents provided informed consent. Students who participated were given a free T-shirt.

    Here will look only at the BMI data for the 94 14 year-old boys in the study. For the boys in this sample, the mean BMI is 23.3, with a standard deviation of 5.4. A study by the Centre for Disease Control[3] at approximately the same time showed that the average BMI for 14 year old boys in the United States was 22.4 kg/m^2. (The CDC study involved a sample, but for the purposes of this question assume that the *true* mean BMI for 14 year-old boys in the United States is 22.4.) Suppose we wish to test whether the study by Duran et al. gives strong evidence that the mean BMI for 14 year-old boys in in rural Texas differs from the BMI for 14 year-old boys in the United States.

    Figure 9.5 illustrates the BMI values for the 94 boys. The red dashed line on the boxplot indicates the hypothesized value of 22.4.



(a) The line represents the hypothesized value of 22.4.

(b) Normal quantile-quantile plot.

Figure 9.5: BMI for 94 boys in rural Texas.

(a) Do these plots show a violation of the normality assumption? If there is a violation, is it a serious problem in this situation? (Can we still use the t procedures?)

[2] Duran et al. (2013). Growth and weight status of rural texas school youth. *American Journal of Human Biology*, 25:71–77. The data used here is simulated data with the same summary statistics as given in their Table 1.

[3] Fryar et al. (2012). Anthropometric reference data for children and adults: United states, 2007–2010. *National Center for Health Statistics. Vital Health Stat.*, 11(252)

(b) In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(c) What is the value of the appropriate test statistic?
(d) What is the p-value of the test?
(e) Give an appropriate conclusion to the hypothesis test.
(f) The output from the statistical software R for this data is:

```
data:  BMI
t = 1.6159, df = 93, p-value = 0.1095
alternative hypothesis: true mean is not equal to 22.4
95 percent confidence interval:
 22.19397 24.40603
sample estimates:
mean of x
     23.3
```

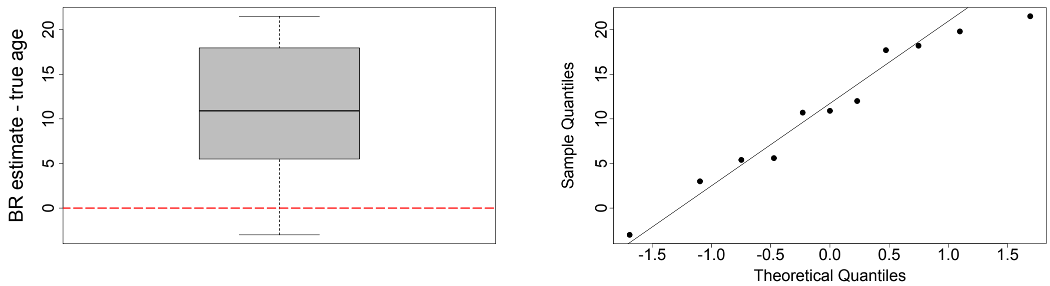
Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply? Comment on any biases that might be present.

63. Accurately estimating the age at death of unidentified human remains is important in forensic science. Is there bias in the procedures used to estimate age at death? A study[4] investigated the validity of 3 dental methods for estimating age at death. In one part of the study, the Bang and Ramm (BR) method was used to estimate the ages of 11 individuals that had a known age of 22. (These were individuals that donated their bodies for science and medical education, or they were from remains in the Weisbach Collection, which is a collection of skeletal material of known age and sex.) The difference between the BR estimate and the true age was recorded (BR estimate − true age), and the average difference was found to be 11.073 years and the standard deviation was found to be 7.784 years. The results are illustrated in Figure 9.6.

If the BR estimation method is unbiased in this situation (the estimated age is correct, on average), then the true (theoretical) mean difference is 0. Does this sample give strong evidence of a bias? Let's investigate this by carrying out a test of the null hypothesis that the true mean difference is 0 against a two-sided alternative hypothesis.

(a) Do these plots show a violation of the normality assumption? Is the use of the t procedures reasonable?

[4] Meinl et al. (2008). Comparison of the validity of three dental methods for the estimation of age at death. *Forensic Science International*, 178:96–105. Values in this question are estimated from their Figure 2. Since the bias of the different aging techniques depends on the real age, this study looked at teeth belonging to bodies of various ages at death. In this question we are looking only at individuals with a true age of 22.



(a) The line represents the hypothesized value of 0.

(b) Normal quantile-quantile plot.

Figure 9.6: The difference between the estimated age and true age for 11 22 year-olds.

(b) In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(c) What is the value of the appropriate test statistic?
(d) What is the p-value of the test?
(e) Give an appropriate conclusion to the hypothesis test.
(f) The output from the statistical software R for this data is:

```
data:  age_at_death
t = 4.718, df = 10, p-value = 0.0008189
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  5.843638 16.302362
sample estimates:
mean of x
   11.073
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply? Comment on any sampling biases that might be present.

64. The ratio of the length of the index finger to the length of the ring finger is called the 2D:4D ratio. A study[5] investigated the 2D:4D ratio in students in an introductory biology course at Western Washington University. The distribution of the 2D:4D ratio is different for groups of different ethnic backgrounds, and this study restricted the sample to people of European descent. In one part of the study, a sample of 135 female students had an average 2D:4D ratio on the left hand of 0.994 and a standard deviation of 0.035.

A question of interest is whether the mean 2D:4D ratio differs from 1. (If an

[5]Stevenson et al. (2007). Attention deficit/hyperactivity disorder (ADHD) symptoms and digit ratios in a college sample. *American Journal of Human Biology*, 19:41–50

individual has a 2D:4D ratio of 1, then their index finger and ring finger are of equal length.) Test the null hypothesis that the population mean 2D:4D ratio is 1, against the alternative hypothesis that it differs from 1. (2D:4D ratios are approximately normally distributed, and in addition the sample size is large here, so the t procedure is appropriate.)

(a) In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(b) What is the value of the appropriate test statistic?
(c) What is the p-value of the test?
(d) Give an appropriate conclusion to the hypothesis test at $\alpha = 0.05$.
(e) The output from the statistical software R for this data is:

```
data:  2D4D
t = -1.9918, df = 134, p-value = 0.04842
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 0.9880422 0.9999578
sample estimates:
mean of x
    0.994
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply? Comment on any biases that might be present.

65. The nutrition information published by a major fast-food chain claims that their quarter-pound cheeseburgers contain 30 grams of protein. A sample of 6 of these cheeseburgers was analyzed, and it was found that they contained an average of 33.81 g of protein with a standard deviation of 2.72 g. (This information is based on sample data published in the U.S. National Nutrient Database.) Suppose we wish to test the null hypothesis that the published nutrition information is correct, and that these cheeseburgers contain 30 g of protein on average. (The use of the t procedures is always a little questionable for such a small sample size ($n = 6$), but there were no outliers in the data so the use the t procedure is not completely out of line.)

(a) In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(b) What is the value of the appropriate test statistic?
(c) What is the p-value of the test?
(d) Give an appropriate conclusion to the hypothesis test at $\alpha = 0.05$.
(e) What assumptions are necessary in order for the test to be valid? What biases might be present?

66. A study[6] investigated concentrations of cadmium and lead in edible mushrooms in an area near the largest Slovenian thermal power plant. Slovenian regulations give a tolerable lead value of 5 mg/kg dw (a lead level of 5 mg/kg of dry weight or less is considered acceptable for consumption). Suppose that the researchers suspected that the lead concentrations in mushrooms in this area might exceed this level, and they wanted to see if they had strong evidence of this. In one aspect of the study, a sample of caps ($n = 21$) from the mushroom *Laccaria amethystina* (commonly known as the Amethyst Deceiver) revealed a mean lead concentration of 0.94 mg/kg dw and a standard deviation of 0.25 mg/kg dw. Does this sample provide strong evidence that the mean lead concentration in this type of mushroom in this area exceeds the tolerable value? Suppose we wish to carry out a t test of the appropriate hypothesis.

    (a) Give the appropriate hypotheses in words and symbols.
    (b) What is the value of the appropriate test statistic?
    (c) What is the p-value of the test?
    (d) Give an appropriate conclusion to the hypothesis test.

67. Franklin et al. (2012)[7] investigated various aspects of tandem running in ants. Tandem running is a form of recruitment in which one ant with knowledge of the location of a food source or new nest site leads another ant to that location. (Optional: Watch an example of tandem running here: (1:53) (http://www.youtube.com/watch?v=X2C7Sy2oPik)) Tandem running can be thought of as a form of teaching and learning.

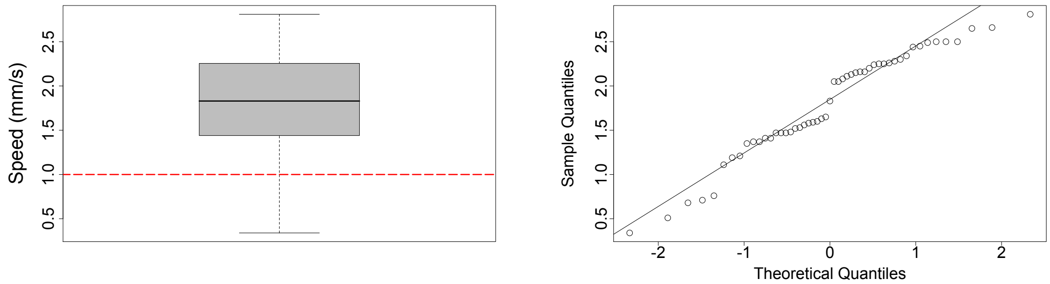
    The study investigated various factors associated with tandem running, and investigated whether the age of the ant and the experience of the ant had any effect on tandem running. Ants (*Temnothorax albipennis*) were categorized into 4 categories: Young and inexperienced (YI), young and experienced (YE), old and inexperienced (OI), old and experienced (OE). One aspect of the study investigated the speed of tandem running for 51 tandem runs in which a YE ant was leading a YI ant. The mean speed (mm/s) of the run was recorded. Figure 9.7 illustrates the speed of the runs. For these 51 runs the sample mean was 1.80 mm/s, and the sample standard deviation was 0.60 mm/s.

    Suppose it was previously believed that under the conditions of this study, the true mean speed of the tandem running would be 1.0 mm/s. Does this study yield strong evidence against this claim?

    (a) Do these plots show a violation of the normality assumption? If there is a

[6]Petkovšek, S. and Pokorny, B. (2013). Lead and cadmium in mushrooms from the vicinity of two large emission sources in Slovenia. *Science of the Total Environment*, 443:944–954

[7]Franklin et al. (2012). Do ants need to be old and experienced to teach? *The Journal of Experimental Biology*, 215:1287–1292 The data used here is simulated data based on their Figure 2, with similar results and conclusions.



(a) The line represents the hypothesized value of 1.0.

(b) Normal quantile-quantile plot.

Figure 9.7: Speed of 51 tandem runs.

violation, is it a serious problem in this situation? (Can we still use the t procedures?)

(b) In words and symbols, what are the hypotheses of the appropriate hypothesis test?

(c) What is the value of the appropriate test statistic?

(d) What is the p-value of the test?

(e) Give an appropriate conclusion to the hypothesis test.

(f) The output from the statistical software R for this data is:

```
data:  Tandem running
t = 9.5219, df = 50, p-value = 8.14e-13
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 1.631247 1.968753
sample estimates:
mean of x
      1.8
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply? Comment on any biases that might be present.

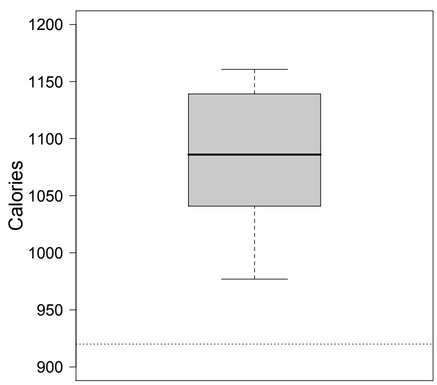
9.12.4 Extra Practice Questions

68. A study by the Center for Science in the Public Interest found that movie theatre popcorn is often very high in calories and saturated fat. The CSPI found that on top of the poor *stated* nutritional characteristics, in reality the popcorn had much higher calorie and fat content than what was claimed.

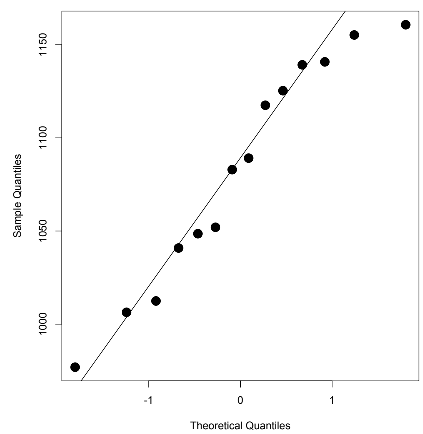
Suppose you wish to investigate the calorie content of large bags of popcorn at your

local theatre, where the theatre chain claims that a large untopped bag of their popcorn contains 920 calories on average. You purchase 14 bags of this type of popcorn, and suppose that it is reasonable to think that these 14 bags represent a random sample from the population of bags of this type. You have these bags of popcorn analyzed and find that the mean calorie content is 1082 and the standard deviation is 60.

A normal quantile-quantile plot and boxplot of the calorie counts are given in Figure 9.8. The theatre's claimed value of 920 calories is represented by a line on the boxplot.



(a) Boxplot of calorie count. The hypothesized value of 920 is indicated by the dotted line.



(b) Normal QQ plot of calorie count.

Figure 9.8: Boxplot and Normal QQ plot of calorie counts.

(a) Do these plots give any indication that the t procedures should not be used here?
(b) Construct a 90% confidence interval for the population mean calorie content of bags of this type.
(c) Suppose we wish to carry out a test of the null hypothesis that the population mean calorie content is what the company claims it to be. If we feel a one-sided alternative hypothesis is most appropriate in this situation, what are the appropriate hypotheses?
(d) Which procedure (z or t) is appropriate here? Why?
(e) What is the value of the appropriate test statistic?
(f) What is the p-value of the test?
(g) Give an appropriate summary of the results of the analysis.

69. The Environment Canada safety guideline for arsenic in soil is 12.0 ppm. (If the

level of arsenic in soil is no more than 12.0 parts per million, the soil is considered safe.) Suppose we suspect that the average arsenic level in the soil in playgrounds of elementary schools in a large school district is greater than 12.0 ppm. We randomly sample 9 schools, take a soil sample from each school, and find that the average arsenic level is 16.7 ppm with a standard deviation of 4.8 ppm. Does this sample provide strong evidence that the true mean arsenic level is greater than 12.0 ppm? (While arsenic levels in soil likely have a right-skewed distribution, for the purposes of this question assume they are normally distributed.)

(a) What are the appropriate hypotheses of the test?
  i. $H_0 : \bar{X} = 12.0, H_a : \bar{X} \neq 12.0$
  ii. $H_0 : \mu = 12.0, H_a : \mu > 12.0$
  iii. $H_0 : \mu = 12.0, H_a : \mu \neq 12.0$
  iv. $H_0 : \mu = 12.0, H_a : \mu < 12.0$
  v. All of the above are appropriate.

(b) Calculate the appropriate test statistic and p-value.

(c) Which one of the following is the most appropriate conclusion at the $\alpha = 0.05$ level of significance?
  i. There is significant evidence that the population mean arsenic level is within Environment Canada guidelines.
  ii. There is significant evidence that the population mean arsenic level is greater than the Environment Canada guideline.
  iii. There is significant evidence that the sample mean arsenic level is within the Environment Canada guideline.
  iv. We know for certain that the population mean arsenic level is greater than the Environment Canada guideline.

(d) Why might this not be the most appropriate test to consider in this situation?

(e) Suppose that the population mean arsenic level is in fact exactly equal to 12.0 ppm. In our hypothesis test, did we make a Type I error, a Type II error, or neither error?

70. Researchers are investigating possible side effects of a newly developed drug. As part of their investigation, the researchers conduct an experiment to see if this drug has an effect on the response time of rats. The researchers know from a great deal of past experience that the average response time for rats that are not given the drug is 1.25 seconds. The researchers administer the drug to a sample of 40 rats and measure the response time. They use the data to test the null hypothesis that the mean response time is 1.25 against a two-sided alternative, and obtain the following output.

```
        One Sample t-test
data:  rat
t = 8.8286, df = 39, p-value = 7.704e-11
alternative hypothesis: true mean is not equal to 1.25
95 percent confidence interval:
 1.424516 1.528247
sample estimates:
mean of x
 1.476382
```

(a) Give a summary of the analysis.
(b) Is it wise to compare the results of their experiment to a historical value? What problems might be associated with this?

71. A researcher is investigating the actual weight of the nuts in packages of nuts that have a stated weight of 250 grams. This researcher buys a sample of packages and weighs the nuts in each package. (Assume that this sample can be thought of as a random sample of bags of this type.) The researcher wishes to carry out an appropriate hypothesis test and obtain a confidence interval for the true mean weight of nuts in bags of this type. The following output summarizes the results of the calculations.

```
        One Sample t-test
data:  nuts
t = 0.9966, df = 44, p-value = 0.3244
alternative hypothesis: true mean is not equal to 250
95 percent confidence interval:
 248.4190 254.6741
sample estimates:
mean of x
 251.5466
```

(a) What is the sample size?
(b) What is the p-value of the test of H_0: $\mu = 250$ against the alternative H_a: $\mu > 250$?
(c) What is the p-value of the test of H_0: $\mu = 250$ against the alternative H_a: $\mu < 250$?
(d) Give a proper interpretation of the results of the hypothesis test found in the output.
(e) Of the following options, which one is the best interpretation of the confidence interval found in the output?
    i. We can be 95% confident that the sample mean is within the interval (248.4190, 254.6741).

ii. 95% of bags of nuts of this type weigh between 248.4190 grams and 254.6741 grams.
iii. We can be 95% confident that the mean weight of all bags of nuts of this type lies between 248.4190 grams and 254.6741 grams.
iv. In repeated sampling, 95% of the sample means will lie within the calculated interval.
v. In repeated sampling, 95% of the sampled bags of nuts will have a weight between 248.4190 grams and 254.6741 grams.

72. In the United States, in order to label a food product as trans fat free, the product must contain no more than 0.50 grams of trans fat per serving. (In Canada a lower threshold of 0.2 g is used.) Suppose that a certain type of french fry in the United States is labelled as trans fat free. A random sample of 150 servings yielded a sample mean trans fat level of 0.538 g, with a sample standard deviation of 0.140 g.

    Test the null hypothesis that the true mean trans fat content is 0.50 g, against the alternative hypothesis that it is greater than 0.50 g.

    (a) Which one of the following represents the appropriate hypotheses in symbols?
        i. H_0: $\mu = 0.50$, H_a: $\mu > 0.50$
        ii. H_0: $\bar{X} = 0.50$, H_a: $\bar{X} > 0.50$
        iii. H_0: $\mu = 0.50$, H_a: $\mu < 0.50$
        iv. H_0: $\bar{X} = 0.528$, H_a: $\bar{X} > 0.528$
        v. Both i and iv are appropriate.
    (b) What is the value of the appropriate test statistic?
    (c) What is the p-value of the test?
    (d) Give an appropriate conclusion at the 0.05 significance level.
    (e) Suppose the distribution of trans fat is strongly skewed (and therefore definitely not normal). Would this hypothesis test still be informative?

73. A manufacturer of bolts is producing a type of bolt that is intended to have an average weight of 82.90 grams. A random sample of 100 of these bolts found that the weights were roughly normally distributed, with a sample mean weight of 84.14 grams and a sample standard deviation of 0.12 grams. Test the null hypothesis that the population mean weight of this type of bolt is 82.9 grams, against the alternative hypothesis that the population mean differs from 82.9 grams.

    (a) Is the appropriate test statistic a t statistic or a z statistic?
    (b) What is the value of the appropriate test statistic?
    (c) What is the p-value of the test?
    (d) Give an appropriate conclusion at the 1% significance level.
    (e) Suppose that the true mean weight of this type of bolt is in fact exactly equal

to 82.90 grams. In our hypothesis test, did we make a Type I error, a Type II error, or neither error?

74. A manufacturer of automobiles has been crash testing cars for many years. In a previous model, it was known from a large body of crash tests that the average driver head injury rating for the crash-test dummies was 800 (units unknown, but the larger the rating, the more damage). This was felt to be unacceptably high, so a new design is implemented and new tests are conducted. Twenty-four cars are crashed and the head injury rating for the drivers was measured. The sample mean was found to be 750 and the sample standard deviation was found to be 50. Test the null hypothesis that the population mean driver head injury for the new design is 800, against the alternative that the new design has a lower population mean head injury rating. (Assume that the head injury ratings are normally distributed.)

    (a) What are the appropriate hypotheses?
    (b) What is the value of the appropriate test statistic?
    (c) What is the p-value of the test?
    (d) Give an appropriate conclusion at the 5% significance level.

75. A car manufacturer is investigating the fuel consumption of a new model of car. As part of the study, 8 of these cars are driven for a certain amount of time under similar real-world conditions and the fuel consumption (litres/100 km) is recorded. The eight cars had a mean fuel consumption of 6.8 l/100k with a standard deviation of 0.32 l/100k. An older model of this car had a mean fuel consumption of 7.0 l/100k, and the manufacturer wishes to test whether the true mean fuel consumption of this new model differs from 7.0 l/100k. Suppose that a normal quantile-quantile plot showed that the fuel consumption values are roughly normally distributed, and that we feel comfortable assuming normality.

    (a) Test the null hypothesis that $\mu = 7$, against a two-sided alternative hypothesis. Give appropriate hypotheses, test statistic, p-value, and conclusion.
    (b) What assumptions are necessary in order for this hypothesis test to be valid? How might these assumptions be violated? What are the consequences if the assumptions are violated?

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. Suppose we draw a random sample from Population 1, which has a mean of 15 and a standard deviation of 2, and we draw a random sample from Population 2, which has a mean of 8 and a standard deviation of 6, and the two samples can be considered independent.

   (a) If $n_1 = n_2 = 4$, what is the mean of the sampling distribution of $\bar{X}_1 - \bar{X}_2$?
   (b) If $n_1 = n_2 = 4$, what is the standard deviation of the sampling distribution of $\bar{X}_1 - \bar{X}_2$?
   (c) If $n_1 = n_2 = 4$, what is the shape of sampling distribution of $\bar{X}_1 - \bar{X}_2$?
   (d) If $n_1 = n_2 = 400$, what is the mean of the sampling distribution of $\bar{X}_1 - \bar{X}_2$?
   (e) If $n_1 = n_2 = 400$, what is the standard deviation of the sampling distribution of $\bar{X}_1 - \bar{X}_2$?
   (f) If $n_1 = n_2 = 400$, what is the shape of sampling distribution of $\bar{X}_1 - \bar{X}_2$?

2. Tom and Pete are two NFL prospects at the NFL combine, where players eligible for the NFL draft show their skills in a variety of different exercises. Tom and Pete are about to have three attempts at the 40 yard dash. Suppose that (theoretically)

Tom's times in this event are approximately normally distributed with a mean of 4.612 seconds and a standard deviation of 0.048 seconds. Pete's times are approximately normally distributed with a mean of 4.528 seconds and a standard deviation of 0.044 seconds. Suppose it is reasonable to assume independence between runs. (The runs may not be truly independent, but this assumption provides a reasonable approximate model.)

(a) In their first attempt, what is the probability that Tom's time is greater than Pete's time?
(b) What is the probability that Tom's average time in the three attempts is greater than Pete's average time in the three attempts?

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. Table 10.1 illustrates the results of a two-sample study. The samples were drawn independently from normally distributed populations.

	Group 1	Group 2
Sample mean	8.8	17.2
Sample standard deviation	1.42	2.61
Sample size	10	5

Table 10.1: Results of a two-sample study.

For the pooled-variance t procedure, calculate:

(a) The point estimate of $\mu_1 - \mu_2$.
(b) The pooled variance s_p^2.
(c) $SE(\bar{X}_1 - \bar{X}_2)$.
(d) A 95% confidence interval for $\mu_1 - \mu_2$.
(e) Test the null hypothesis that the population means are equal. Give the appropriate hypotheses, standard error, value of the test statistic, p-value, and conclusion at $\alpha = 0.05$.

4. Does a vitamin D supplement have an effect on parathyroid hormone (PTH) levels in the blood? An experiment[1] investigated the effect of a vitamin D supplement on several biological factors in study participants. One of the variables was the change in PTH in the blood.

   In the experiment, 26 individuals were randomly assigned to one of two groups. Each group consumed 240 mL of orange juice per day for 12 weeks. The orange juice of the treatment group was fortified with 1000 IU vitamin D_3, whereas the control group's orange juice had no vitamin D_3 added. After 12 weeks, the change in PTH level in the blood (pg/mL) was recorded. Table 10.4 illustrate the results.

Fortified with vitamin D	$\bar{X}_1 = -9.0$	$s_1 = 37.5$	$n_1 = 14$
Not fortified with vitamin D	$\bar{X}_2 = -1.6$	$s_2 = 34.6$	$n_2 = 12$

Table 10.2: Change in blood PTH levels after 12 weeks.

   Suppose that normal quantile-quantile plots showed that the values were approximately normally distributed. Use the pooled-variance t procedure to answer the following questions. (Since the observations are approximately normally distributed, and the sample standard deviations are similar, the pooled-variance t procedure is a reasonable method of analysis.)

   (a) Test the hypothesis that vitamin D has no effect against a two-sided alternative. Give the appropriate hypotheses in words and symbols, value of the test statistic, p-value, and conclusion. (Hint to ease the calculation burden: $s_p^2 = 1310.417$, $SE(\bar{X}_1 - \bar{X}_2) = 14.24088$.)
   (b) Calculate a 95% confidence interval for the difference in population mean PTH level. Give a proper interpretation of the interval.
   (c) Give an overall summary of the results of the analysis.

10.4.2 Welch (Unpooled Variance) t Tests and Confidence Intervals

5. Table 10.3 illustrates the results of a two-sample study. The samples were drawn independently from normally distributed populations.

   For the Welch procedure, calculate:

   (a) The point estimate of $\mu_1 - \mu_2$.
   (b) $SE_W(\bar{X}_1 - \bar{X}_2)$.

[1] Tangpricha et al. (2003). Fortification of orange juice with vitamin D: a novel approach for enhancing vitamin D nutritional health. *The American journal of clinical nutrition*, 77(6): 1478-1483.

	Group 1	Group 2
Sample mean	8.8	17.2
Sample standard deviation	1.42	2.61
Sample size	10	5

Table 10.3: Results of a two-sample study.

(c) A 95% confidence interval for $\mu_1 - \mu_2$. (Hint to ease the calculation burden: DF = 5.221.)

(d) Test the null hypothesis that the population means are equal. Give the appropriate hypotheses, standard error, value of the test statistic, p-value, and a conclusion at $\alpha = 0.05$. (Hint to ease the calculation burden: DF = 5.221.)

6. Consider again the study discussed in Question 4. (In this experiment, some individuals received orange juice fortified with vitamin D, others received unfortified orange juice.) Does the vitamin D supplement affect phosphorous levels in the blood? The change in phosphorous levels over the course of the study are found in the following table.

Fortified with vitamin D	$\bar{X}_1 = -6.3$	$s_1 = 18.7$	$n_1 = 14$
Not fortified with vitamin D	$\bar{X}_2 = -0.2$	$s_2 = 1.7$	$n_2 = 12$

Table 10.4: Change in blood phosphorous levels (mg/dL) after 12 weeks.

Note that the standard deviations are vastly different here (so much so that it makes one wonder if there might be a typo in the article). Since the standard deviations are vastly different, the pooled-variance procedure would start to break down and would not be appropriate. The Welch procedure would be a better method of analysis, so use the Welch procedure to answer the following questions.

(a) What is the value of $SE_W(\bar{X}_1 - \bar{X}_2)$?

(b) If we wish to test the null hypothesis that, on average, vitamin D has no effect on phosphorous levels, what is the value of the appropriate test statistic?

(c) The R output for the Welch procedure is:

```
        Welch Two Sample t-test
data:  vitaminD and novitaminD
t = -1.2156, df = 13.215, p-value = 0.2454
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -16.923691   4.723691
sample estimates:
mean of x mean of y
     -6.3      -0.2
```

Give a summary of the results of the analysis.

10.4.3 Guidelines for Choosing the Appropriate Two-Sample t Procedure

7. What factors influence the choice between the Welch procedure and the pooled-variance t procedure? When is the Welch procedure a better choice? When is the pooled-variance t procedure a better choice?

8. Suppose we are interested in assessing the effect of two fuel additives on fuel efficiency. We intend to use a t procedure to carry out an appropriate hypothesis test. Under which of the following situations would it be most appropriate to use Welch's approximation instead of the pooled-variance procedure?

   (a) The sample sizes of the two groups are similar, and the sample standard deviations are similar.
   (b) The sample sizes of the two groups are similar, and the sample standard deviations are very different.
   (c) The sample sizes of the two groups are very different, and the sample standard deviations are similar.
   (d) The sample sizes of the two groups are very different, and the sample standard deviations are very different.

10.5 Paired-Difference Procedures

10.5.1 Paired-Difference t Tests and Confidence Intervals

9. Offerman et al. (2009) conducted an experiment on pigs to investigate the effect of an antivenom after an injection of rattlesnake venom. In one aspect of the study, the researchers investigated the change in volume of the right hind leg before and

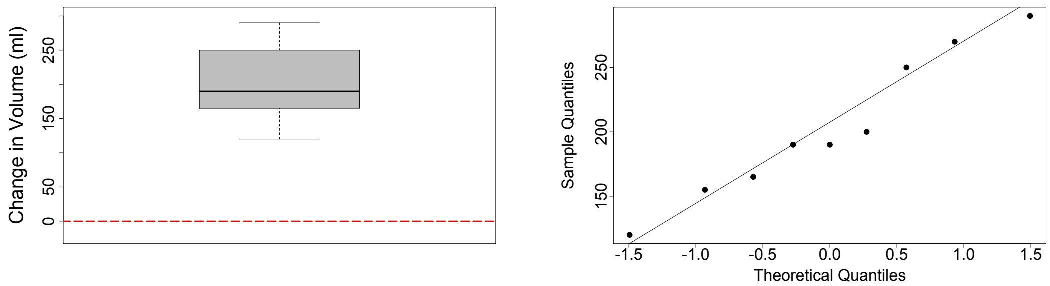
after being subjected to a dose of venom and treatment with an antivenom. The volume of the right hind leg was measured in 9 pigs before being injected with the venom, then the pigs were injected with a dose of venom and treated intravenously with an antivenom, and after 8 hours the volume of the leg was measured again. The results are illustrated in Table 10.5. (The volume was measured using a water displacement method; the units are mL.)

Before	After	Difference
685	935	
545	700	
480	770	
475	640	
680	800	
685	955	
590	780	
600	790	
630	830	

Table 10.5: Volume (mL) of the right hind leg before and 8 hours after injection with a dose of rattlesnake venom and an antivenom.

(a) Suppose we are interested in estimating the true mean amount of swelling for pigs of this type under the conditions of this experiment. Would we use a paired-difference procedure or an independent sample procedure? Justify your response.
(b) What are the 9 differences? (Take the differences as After − Before.)
(c) What is the sample mean difference? What is the standard deviation of the differences?
(d) What is the standard error of the sample mean difference?
(e) The boxplot and normal quantile-quantile plot of the differences are given in Figure 10.1. Do these plots give any indication that the t procedures should not be used?
(f) Construct a 95% confidence interval for the true mean amount of swelling under the conditions of this experiment. Give an interpretation of the confidence interval in the context of the problem at hand.
(g) Carry out a test of the null hypothesis that there is no swelling on average (after 8 hours), against the appropriate one-sided alternative hypothesis. Give the hypotheses in words and symbols, value of the test statistic, p-value, and conclusion.

10. Porter et al. (2010) investigated the effect of a brief training program on the ability of health care professionals to detect deception. In part of the study, 26 health care workers were shown videos in which individuals sometimes showed a genuine



(a) The line represents the hypothesized value of 0.

(b) Normal quantile-quantile plot.

Figure 10.1: Changes in volume (mL) of the right hind leg before and 8 hours after injection with a dose of rattlesnake venom and an antivenom.

smile, and sometimes showed a fake smile. Individuals completed this task before and after completing a 3 hour session designed to help them recognize deception. The participants' responses were evaluated and they were given a *discrimination accuracy* score (high positive scores indicate a person is correctly discriminating between a genuine smile and a fake smile, values near 0 indicate the individual is not discriminating between a genuine smile and a fake smile, and negative values indicate that the individual is misclassifying genuine smiles as fake and fake smiles as genuine). The results of the study are given in Table 10.6.

	Mean	SD
Before	0.192	0.770
After	0.810	1.173
Difference (After − Before)	0.618	1.412

Table 10.6: Discrimination accuracy scores for 26 health care professionals before and after a training program.

The researchers were interested in testing whether the training program had an effect, and in estimating the size of the effect with a confidence interval.

(a) What should be done before using software to carry out the t procedures?

(b) When the calculations are carried out in R, the default output is:

```
        One Sample t-test
data:  expressions
t = 2.2317, df = 25, p-value = 0.0348
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.04768086 1.18831914
sample estimates:
mean of x
    0.618
```

The p-value given in the output iss for a two-sided alternative hypothesis. But suppose that the researchers felt a one-sided alternative hypothesis was more appropriate, since they felt that, if anything, the training session would improve a participant's ability to discriminate between genuine and fakes smiles. What would be the appropriate one-sided alternative? What would the p-value be for this test?

(c) Give a summary of the results of the analysis. Use a significance level of 0.05.

10.6 Investigating the Normality Assumption

10.7 Chapter Exercises

10.7.1 Basic Calculations

10.7.2 Concepts

11. In words, what is the meaning of $SE(\bar{X}_1 - \bar{X}_2)$?

12. Would it make sense to test the hypothesis H_0: $\bar{X}_1 = \bar{X}_2$? Why or why not?

13. A 95% confidence interval for $\mu_1 - \mu_2$ is found to be $(2, 28)$.

    (a) Give an example of a null hypothesis that would be rejected at $\alpha = 0.05$.
    (b) Give an example of a null hypothesis that would not be rejected at $\alpha = 0.05$.

14. Consider the following output for a two-sample problem.

```
        Two Sample t-test
t = -0.5692, df = 18, p-value = 0.5763
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9817025  0.5631646
sample estimates:
  mean of x   mean of y
-0.12739020  0.08187872
```

Does this output yield strong evidence that $\bar{X}_1$ does not equal $\bar{X}_2$?

15. Suppose we wish to test the null hypothesis that $\mu_1 - \mu_2 = 0$ against a two-sided alternative hypothesis.

    (a) All else being equal, what will happen to the power of the test as the true difference $\mu_1 - \mu_2$ gets closer to 0?
    (b) All else being equal, what will happen to the power of the test as the sample sizes increase?

16. Consider a two-sample t test of the null hypothesis of equal population means against a two-sided alternative.

    (a) Under what conditions would the test statistic be equal to 0?
    (b) Under what conditions would the p-value be equal to 1?
    (c) Under what conditions would the p-value be equal to 0?

17. Consider the three boxplots in Figure 10.2, which represent samples of size 40 from 3 different populations. Consider the following 3 null hypotheses (with two-sided

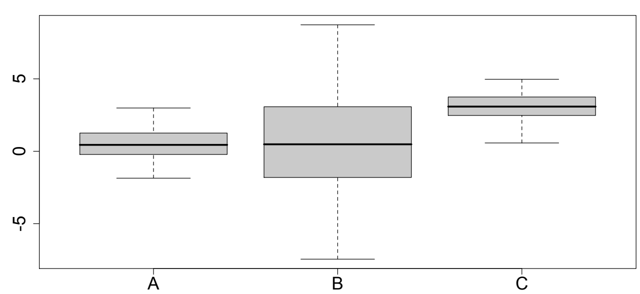


Figure 10.2

alternatives in all cases).

    I. H_0: $\mu_A = \mu_B$
    II. H_0: $\mu_A = \mu_C$

III. H_0: $\mu_B = \mu_C$

(a) Which test would result in the smallest p-value?
(b) Which test would result in the largest p-value?

18. Consider the following output for a two-sample inference procedure, and the corresponding boxplots in Figure 10.3.

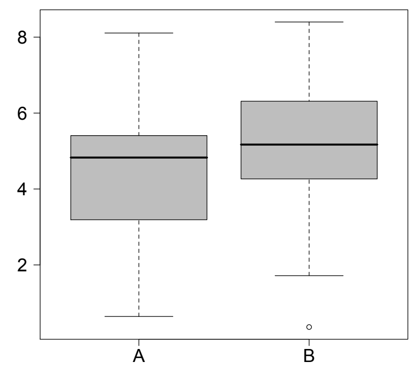


Figure 10.3

```
        Two Sample t-test
t = -1.0655, df = 43, p-value = 0.2926
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.6882973  0.5209971
sample estimates:
mean of x mean of y
 4.367041  4.950691
```

(a) Do the boxplots give any indication that this t procedure should not be used?
(b) Is there strong evidence that the populations have different means?
(c) Based on the confidence interval in the output, is there strong evidence that $\mu_1 - \mu_2 \neq 10$?
(d) What is the sum of the two sample sizes?

19. Suppose we draw two independent samples of sizes $n_1 = 100$ and $n_2 = 50$, and wish to test the null hypothesis that the population means are equal. The output for the two procedures (Welch and the pooled-variance t) are:

Output 1:

```
data:  sample1 and sample2
t = -1.3734, df = 98.659, p-value = 0.1727
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5677384  0.1032975
sample estimates:
 mean of x  mean of y
-0.1298456  0.1023749
```

Output 2:

```
data: sample1 and sample2
t = -1.3705, df = 148, p-value = 0.1726
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5670558  0.1026149
sample estimates:
 mean of x  mean of y
-0.1298456  0.1023749
```

(a) Which output is from the pooled-variance procedure? Which output is from the Welch procedure?
(b) Which procedure would be more appropriate here?

20. Forty patients with high blood pressure volunteer for a study. The participants are given an injection of a placebo (a saline solution with no pharmacological effect), and an hour later their drop in blood pressure is measured. Two days later the same 40 patients are given an injection of an experimental new drug, and an hour later the drop in blood pressure is recorded. Researchers want to investigate the effect of the drug on blood pressure. They wish to compare the drop in blood pressure after the injection of the drug to the drop in blood pressure after injection of the placebo.

    (a) What t procedure would be the most appropriate procedure to use here (Welch's, pooled-variance, or paired difference)? Suppose that it is reasonable to assume normality where necessary.
    (b) Suppose we observe a significant difference between the drop in blood pressure observed after the placebo injection, and the drop in blood pressure after the drug injection. Other than a possible effect of the drug, are there any features of this study design that may possibly be the cause of this difference?

21. A research experiment was designed to assess the effectiveness of two memory training programs. The experiment involved 25 sets of identical twins. One twin in each pair was randomly assigned to Program A, the other to Program B. Upon completion of the program, everyone took a memory test and obtained a score between 0 and 100. If we want to carry out a hypothesis test to see if one of the programs

tends to result in a better score on the memory test, what is the most appropriate test to use? Assume normality where necessary.

22. (Challenge! We haven't worked through a problem like this—you need to think it through.) Suppose we wish to draw independent random samples from two populations. Suppose that it is known that $\sigma_1 = \sigma_2 = 3$, but the population means are unknown. We intend to draw equal sample sizes $n_1 = n_2 = n$ from both populations. If we wish to estimate $\mu_1 - \mu_2$ within 0.5 with 95% confidence, what sample size would be required?

23. 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) When we use the pooled-variance t procedure, it is because we *know* the populations have the same variance.
    (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.
    (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.
    (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.
    (e) The Welch procedure is an exact procedure, as long as $\bar{X}_1 = \bar{X}_2$.

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) If $\mu_1 = \mu_2$, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is approximately symmetric about 0 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.
    (c) $SE(\bar{X}_1 - \bar{X}_2)$ is the true standard deviation of the sampling distribution of $\bar{X}_1 - \bar{X}_2$.
    (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.
    (e) Suppose we about to test 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.

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) 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.

(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.

(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.

(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.

(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 $\mu_1 = \mu_2$.

10.7.3 Applications

26. A study[2] investigated various aspects of tandem running in ants. Tandem running is a form of recruitment in which one ant with knowledge of the location of a food source or new nest site leads another ant to that location. (Optional: Watch an example of tandem running here: (1:53) (http://www.youtube.com/watch?v=X2C7Sy2oPik)) Tandem running can be thought of as a form of teaching and learning.

The study investigated various factors associated with tandem running, and investigated whether the age of the ant and the experience of the ant had any effect on tandem running. Ants (*Temnothorax albipennis*) were categorized into 4 categories: Young and Inexperienced (YI), Young and Experienced (YE), Old and Inexperienced (OI), Old and Experienced (OE). One aspect of the study investigated the speed of tandem running when leading a YI ant. The mean speed of the run was recorded for 51 tandem runs in which the leader was YE, and 15 pairs of runs in which the leader was OE. The researchers were interested in a possible difference in the mean speed of tandem running in these two situations.

Figures 10.4 and 10.5 and Table 10.7 illustrate the data.

OE	$\bar{X}_1 = 1.14$	$s_1 = 0.37$	$n_1 = 15$
YE	$\bar{X}_2 = 1.80$	$s_2 = 0.60$	$n_2 = 51$

Table 10.7: Means and standard deviations for the tandem running study.

(a) Do the plots give any indication that the t procedure should not be used?

[2] Franklin et al. (2012). Do ants need to be old and experienced to teach? *The Journal of Experimental Biology*, 215:1287–1292 The data used here is simulated data based on their Figure 2, with similar results and conclusions.

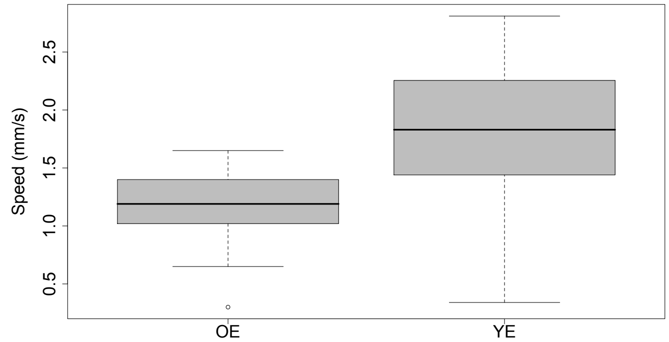
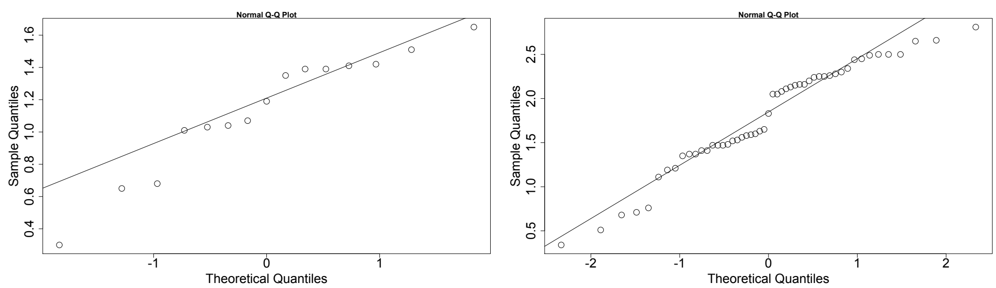


Figure 10.4: Speed of tandem running for ant pairs led by an old experienced (OE) ant, and ant paris led by a young experienced (YE) ant.



(a) OE ants.

(b) YE ants.

Figure 10.5: Normal QQ plots for OE and YE ants.

(b) Which version of the t procedure (pooled or unpooled) is more appropriate here?

(c) Using the Welch procedure, construct a 95% confidence interval for the difference in true mean tandem running speed between OE and YE ants. (Hint to ease the calculation burden: $SE_W(\bar{X}_1 - \bar{X}_2) = 0.1272$, DF $= 37.714$.)

(d) The researchers were interested in investigating a possible difference in the true mean running speeds for the two types of ant. In words and symbols, what are the hypotheses of the appropriate hypothesis test?

(e) What is the value of the appropriate test statistic?

(f) What is the p-value of the test? Is there strong evidence against the null hypothesis?

(g) The output from the statistical software R for the Welch procedure is:

```
        Welch Two Sample t-test
data:  OE and YE
t = -5.1878, df = 37.714, p-value = 7.516e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9176121 -0.4023879
sample estimates:
mean of x mean of y
     1.14      1.80
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply?

27. A study[3] investigated several aspects of the ratio of the lengths of the index finger to the ring finger in women who visited a sexual health clinic in Manchester, UK. (The ratio of the lengths of these fingers is called the 2D:4D ratio.) The distribution of the 2D:4D ratio depends on several factors, including the ethnic background of the individuals. The following table summarizes the results of the measurements of the 2D:4D ratio on the left hand for white and black women in the study.

White women	$\bar{X}_1 = 0.994$	$s_1 = 0.035$	$n_1 = 246$
Black women	$\bar{X}_2 = 0.963$	$s_2 = 0.034$	$n_2 = 46$

(a) What plots should be created before carrying out any inference procedures?
(b) Which version of the t procedure (pooled or unpooled) is more appropriate here?
(c) Using the pooled-variance t procedure, construct a 95% confidence interval for the difference in the means of the 2D:4D ratio. (Hint to ease the calculation burden: $SE(\bar{X}_1 - \bar{X}_2) = 0.005598$.)
(d) Suppose we wish to test whether there is strong evidence of a difference between the true means of the 2D:4D ratio for white and black women. In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(e) What is the value of the appropriate test statistic?
(f) What is the p-value of the test? Is there strong evidence against the null hypothesis?
(g) The output from the statistical software R for the pooled-variance procedure is:

[3]Brabin et al. (2008). The second to fourth digit ratio (2d:4d) in women with and without human papillomavirus and cervical dysplasia. *American Journal of Human Biology*, 20:337–341.

```
        Two Sample t-test
data:  white and black
t = 5.538, df = 290, p-value = 6.859e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01998280 0.04201720
sample estimates:
mean of x mean of y
    0.994     0.963
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply?

28. Researchers used an experiment to investigate whether prolonged food restriction might lead to a relapse of drug abuse.[4] Rats were randomly assigned to a sated (well fed) group or a prolonged food restriction group (the rats were under food restriction for 10 days, and their weight fell to approximately 80% of the weight of the sated rats). Before the food deprivation period, all rats were trained to use a lever to self-administer heroin. Access to heroin was removed for all rats during the 10 day food deprivation period. Rats were then given access to the heroin lever, and the number of active lever presses in a 3 hour period was measured. The following table summarizes the results.

Food restriction	$\bar{X}_1 = 48.9$	$s_1 = 25.4$	$n_1 = 8$
Sated (control)	$\bar{X}_2 = 23.8$	$s_2 = 7.9$	$n_2 = 7$

(a) What plots should be created before carrying out any statistical inference?

(b) Plots of the data (not shown) show some right skewness in both groups, and so the use of the t procedures is a bit dubious. But if the skewness is similar in both groups, the t procedure may still perform reasonably well. Assuming that we choose to use the t procedures, which version of the t procedure (pooled or unpooled) is more appropriate here?

(c) Using the Welch (unpooled variance) t procedure, construct a 95% confidence interval for the difference in the means between the groups. (Hint to ease the calculation burden: $SE_W(\bar{X}_1 - \bar{X}_2) = 9.4637$, DF = 8.512.)

(d) Suppose we wish to test whether there is strong evidence of a treatment effect (a difference between the true means of the two groups). In words and symbols, what are the hypotheses of the appropriate hypothesis test?

(e) What is the value of the appropriate test statistic?

(f) What is the p-value of the test? Is there strong evidence against the null hypothesis?

[4]Shalev (2011). Chronic food restriction augments the reinstatement of extinguished heroin-seeking behavior in rats. *Addiction Biology*, 17:691–693.

(g) The output from the statistical software R for the pooled-variance procedure is:

```
        Welch Two Sample t-test
data: Restricted and Sated
t = 2.6523, df = 8.512, p-value = 0.02764
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  3.50315 46.69685
sample estimates:
mean of x mean of y
     48.9      23.8
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply?

29. A study[5] investigated several physical characteristics of the ears of Italian Caucasians. In one part of the study, 3D symmetry (a measure of the symmetry between the left and right ears) was measured on men and women between 31 and 40 years of age.

Men	$\bar{X}_1 = 95.71$	$s_1 = 1.50$	$n_1 = 66$
Women	$\bar{X}_2 = 95.41$	$s_2 = 1.64$	$n_2 = 28$

(a) What plots should be created before carrying out any inference procedures?
(b) Which version of the t procedure (pooled or unpooled) is more appropriate here?
(c) Using the pooled-variance t procedure, construct a 95% confidence interval for the difference in the means of the 3D symmetry index. (Hint to ease the calculation burden: $SE(\bar{X}_1 - \bar{X}_2) = 0.3479$.)
(d) Suppose we wish to test whether there is strong evidence of a difference between the true mean 3D symmetry in men and women. In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(e) What is the value of the appropriate test statistic?
(f) What is the p-value of the test? Is there strong evidence against the null hypothesis?
(g) The output from the statistical software R for the pooled-variance procedure is:

[5]Sforza et al. (2009). Age- and sex-related changes in the normal human ear. *Forensic Science International*, 187:110.e1–110.e7.

```
        Two Sample t-test
data: men and women
t = 0.8624, df = 92, p-value = 0.3907
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3908909  0.9908909
sample estimates:
mean of x mean of y
    95.71     95.41
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply?

30. A study[6] investigated several characteristics of psychopathic and nonpsychopathic male patients in a Dutch inpatient psychiatric treatment centre. In one part of the study, the BEST index (a measure of risk behaviours of psychiatric patients) was measured after several months of treatment. The higher the score on the BEST index, the better (less risky) the patient scored. The following table summarizes the results.

Nonpsychopaths	$\bar{X}_1 = 256.12$	$s_1 = 33.53$	$n_1 = 47$
Psychopaths	$\bar{X}_2 = 245.82$	$s_2 = 37.55$	$n_2 = 27$

(a) What plots should be created before carrying out any inference procedures?
(b) Which version of the t procedure (pooled or unpooled) is more appropriate here?
(c) Using the pooled-variance t procedure, construct a 95% confidence interval for the difference in the means of the BEST index. (Hint to ease the calculation burden: $SE(\bar{X}_1 - \bar{X}_2) = 8.4603$.)
(d) Suppose we wish to test whether there is strong evidence of a difference between the true means of the BEST index for psychopaths and nonpsychopaths. In words and symbols, what are the hypotheses of the appropriate hypothesis test?
(e) What is the value of the appropriate test statistic?
(f) What is the p-value of the test? Is there strong evidence against the null hypothesis?
(g) The output from the statistical software R for the pooled-variance procedure is:

[6]Chakhssi et al. (2010). Change during forensic treatment in psychopathic versus nonpsychopathic offenders. 21:660–682.

```
        Two Sample t-test
data: psychopaths and non psychopaths
t = 1.2174, df = 72, p-value = 0.2274
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.565314 27.165314
sample estimates:
mean of x mean of y
   256.12    245.82
```

Give a summary of the results of the analysis (including the results of the hypothesis test and the confidence interval). To what population do your conclusions apply?

31. What effect does a shock with a Taser have on the components of blood? Jauchem et al. (2013) investigated this in an experiment on 11 wild boar (*Sus scrofa*) by measuring characteristics of blood before and after a Taser shock. One measured characteristic was the mean corpuscular volume (the average volume of red blood cells, measured in femtolitres (10^{-15} L)).

	Mean	SD
Before	54.5	2.4
After	57.8	2.6
Difference (After – Before)	3.3	3.7

Table 10.8: Summary statistics of the mean corpuscular volume (fL) for 11 wild boar before and after a Taser shock.

Suppose that the changes in mean corpuscular volume are approximately normally distributed.

(a) Construct a 95% confidence interval for the true mean change in mean corpuscular volume, and give an interpretation of the interval.
(b) Carry out a test of the null hypothesis that, on average, there is no change in mean corpuscular volume, against a two-sided alternative hypothesis. Give the hypotheses in words and symbols, value of the test statistic, p-value, and conclusion.

10.7.4 Extra Practice Questions

32. In an observational study involving 42 new mothers, the women were classified according to whether or not they had used marijuana in their pregnancy. The point of interest was comparing the birth weight of the babies born to the two groups.

Did not use marijuana	$\bar{X}_1 = 3468$	$s_1 = 680$	$n_1 = 32$
Used marijuana	$\bar{X}_2 = 3200$	$s_2 = 610$	$n_2 = 10$

The researchers wanted to use a reasonable statistical analysis to investigate if there is a difference in mean birth weight between the two groups of mothers. The researchers felt it was reasonable to assume normality and to assume the two population variances are equal. Using the appropriate procedure under these conditions, they researchers found an output of:

```
Two Sample t-test
data: marijuana and nomarijuana
t = 1.1126, df = 40, p-value = 0.2725
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
****** ******
sample estimates:
mean of x mean of y
     3468      3200
```

(a) The 90% confidence interval has been omitted from the output. What is the appropriate interval? (Hint to ease the calculation burden: $SE(\bar{X}_1 - \bar{X}_2) = 240.8803$.)

(b) Which one of the following is the most appropriate conclusion?

i. There is very strong evidence that the population mean weight of babies born to mothers who used marijuana is equal to that of babies born to mothers who did not use marijuana.
ii. There is not strong evidence of a difference in population mean birth weight between the two groups.
iii. There is strong evidence that the population mean birth weights for the two groups are equal.
iv. There is not strong evidence that the sample mean birth weights are different.
v. There is very strong evidence that marijuana causes a reduction in birth weight.

33. An experiment was designed to investigate the effect of exercise on the size of tumours in rats. Thirty rats were injected with cancerous cells. These rats were then randomly assigned to two groups. Ten mice were kept in cages with exercise wheels, and 20 were kept in cages with no wheels. After six weeks, the diameter of the tumour (cm) was recorded. The following table summarizes the results:

	Exercise Wheels	No Wheels
Sample mean	1.2	1.5
Sample variance	0.08	0.12

(a) Is this an observational study or an experiment?

(b) The researcher wanted to test whether the exercise wheels reduced the size of tumours in rats. Carry out the appropriate hypothesis test. Give appropriate hypotheses, the value of the test statistic, p-value, and an appropriate conclusion. (Assume equal population variances, and that the tumour sizes are normally distributed.) (Hint to ease the calculation burden: $s_p^2 = 0.10714$, $SE(\bar{X}_1 - \bar{X}_2) = 0.12677$)

(c) Suppose we found a p-value of .00000067 in the previous question. Would this give strong evidence of a causal link between exercise and size of tumour in this type of experiment?

(d) Calculate a 90% confidence interval for the difference in mean tumour size (for the pooled variance case).

(e) Now, *not* assuming equal population variances, carry out the appropriate test. Give appropriate hypotheses, value of the test statistic, p-value, and an appropriate conclusion. (Hint to ease the calculation burden: DF = 21.764.)

(f) Calculate a 90% confidence interval for the difference in mean tumour size. (Not assuming equal population variances.) (Hint to ease the calculation burden: DF = 21.764.)

(g) Which procedure (pooled-variance or Welch) is more appropriate in this case?

34. Researchers investigated the total cholesterol levels in the blood of male and female students at a large university. Total cholesterol (mg/dl) was measured on 26 male and 22 female student volunteers, with the following results.

	Males	Females
Sample mean	171.4	173.8
Sample standard deviation	32.9	34.1
Sample size	26	22

Although the people in the study are volunteers, for the purposes of these questions assume they can be thought of as random samples from the populations.

(a) The researchers wanted to test the null hypothesis that the population mean total cholesterol level is the same for both males and females, against a two-sided alternative hypothesis. The following output represents the results of the pooled variance two-sample t procedure.

```
Two Sample t-test
data:  males and females
t = -0.2477, df = 46, p-value = 0.8055
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21.90658  17.10658
sample estimates:
mean of x mean of y
    171.4     173.8
```

Of the following options, which one is the most appropriate conclusion at $\alpha = .05$?

i. There is strong evidence of a difference in population means.
ii. There is significant evidence that females have a higher population mean total cholesterol level.
iii. The observed difference in sample mean cholesterol levels is significant.
iv. There is not significant evidence that the sample mean cholesterol levels are different.
v. There is not significant evidence that the population mean cholesterol levels are different.

(b) Give an appropriate interpretation of the confidence interval found in the output.

35. In an investigation into the effectiveness of two processes that reduce contaminants in used motor oil, 18 batches of used motor oil were randomly assigned to the two processes. 12 were randomly assigned to process A, and 6 to process B. The contaminant level after processing was measured for both processes. The results are given in Table 10.9.

	A	B
Sample size	12	6
Sample variance	16	9
Sample mean	63	57

Table 10.9: Summary of information for the motor oil question.

Suppose that the normality assumption of the t procedures is in fact reasonable, and that the population variance of the contaminants is the same for both processing methods.

(a) Calculate a 95% confidence interval for the difference in the population mean contaminant levels ($\mu_A - \mu_B$). Give an appropriate interpretation of the interval.

(b) Carry out a test of the null hypothesis that the true mean contaminant level is the same for both processes, against the alternative hypothesis that it is different. Give appropriate hypotheses in words and symbols, value of the test statistic, and p-value.

(c) Summarize the results of the analysis.

36. Many studies have investigated a connection between "fear of negative evaluation" and bulimia. Suppose that researchers at a large university are interested in carrying out their own investigation. Eleven female students with bulimia completed a questionnaire and were assigned a "fear of negative evaluation" score, with a resulting sample mean of 19.7. Fourteen female students with normal eating habits were given the same questionnaire, with a resulting sample mean "fear of negative evaluation" score of 14.9. The researchers ran a pooled-variance two-sample t procedure on the data, with the following results.

```
Two Sample t-test
data:  bulimic and notbulimic
t = 2.5621, df = 23, p-value = 0.01742
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.9380235 8.8022362
sample estimates:
mean of x mean of y
 19.72727  14.85714
```

Assume that the students can be thought of as random samples from the populations of female students at the university, and the other assumptions of the pooled-variance t-procedures are reasonable in this case.

(a) In words and symbols, what are the hypotheses of the t test given in the output?

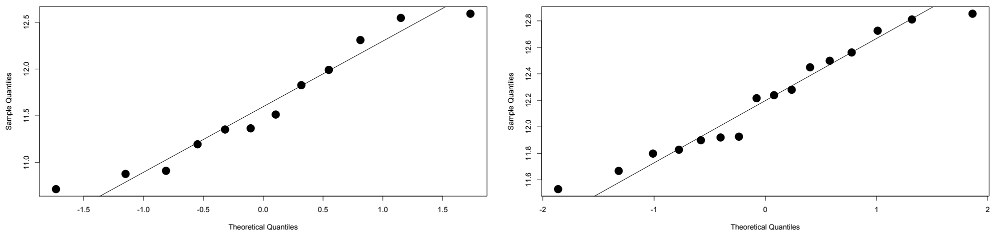
(b) Give an appropriate conclusion at $\alpha = .05$.

(c) Give a proper interpretation of the confidence interval found in the output.

37. As part of a production process, a company needs a certain type of resistor to have a resistance of 12 ohms. The company obtains resistors from 2 different suppliers and tests their resistance. The results are given in the table below. Although it would not be perfectly justified in this case, assume the resistors can be thought of an independent random samples from their respective populations.

Manufacturer	Sample size	Sample mean	Sample standard deviation
Manufacturer A	12	11.60	0.65
Manufacturer B	16	12.20	0.42

The researchers want to investigate several properties of these resistors. As part of



(a) Manufacturer A (b) Manufacturer B

Figure 10.6: Normal quantile-quantile plots of the resistance.

their first investigation, they want to estimate the difference in the mean resistance between the two types of resistor, and test the null hypothesis that the resistors have the same resistance on average. They wish to use the t procedures, but realize that they should first investigate the normality assumption. Figure 10.6 illustrates the normal quantile-quantile plots in this scenario.

(a) Do these plots give any indication that the normality assumption is violated?

(b) As part of their analysis into the differences between the resistors, the researchers want to test the null hypothesis that the population mean resistance is the same for both types of resistor, against the alternative hypothesis that the mean resistance is different. They also wish to calculate a 95% confidence interval for the difference in the mean resistance. They decide to use the **pooled-variance t procedure**. The Welch procedure would have also been a reasonable choice (the output for the Welch procedure is included farther below). What is the value of the pooled sample variance (s_p^2)?

(c) What are the appropriate hypotheses in words and symbols?

(d) What is the value of the appropriate t statistic?

(e) Give an appropriate conclusion at the 5% significance level.

(f) What is a 95% confidence interval calculated using the pooled-variance t procedure?

(g) The following output represents the results of the Welch procedure on the data above, using the alternative hypothesis that the population mean resistance for the two types of resistors is different.

```
Welch Two Sample t-test
data:  resistors
t = -2.7904, df = 17.695, p-value = 0.01222
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.0522974 -0.1477026
```

```
sample estimates:
mean of x mean of y
     11.6      12.2
```

Based on the confidence interval in the output, what can be said about possible differences in the mean resistance?

(h) Refer again to the output from the Welch procedure. Note that the output gives the p-value for a two-sided alternative. Had the researchers felt that H_a: $\mu_A < \mu_B$ was the more appropriate alternative (before looking at the data), what would the p-value of that test be?

38. A company is interested in investigating properties of a new aluminum alloy that is produced by a new experimental process. They wish to compare the yield strength of this new alloy to the yield strength of a standard alloy that is commonly used. The following table gives a summary of the yield strength (MPa) for independent samples of each type of alloy.

Alloy type	Sample size	Sample mean	Sample standard deviation
New alloy	80	664.2	23.3
Standard alloy	75	624.3	19.8

Suppose it is reasonable to assume that the alloy strengths are approximately normally distributed, and we wish to carry out inference procedures to investigate possible differences between the alloys.

There is little difference between the sample standard deviations, so the use of the pooled-variance t procedure is reasonable here. Use the pooled-variance t procedure to answer the following questions.

(a) Calculate a 95% confidence interval for the difference in true mean yield strength. Give a proper interpretation of the interval. (Hint to ease the calculation burden: $s_p^2 = 469.9299$, $SE(\bar{X}_1 - \bar{X}_2) = 3.4842$.)

(b) Carry out a test of the null hypothesis that the population means are equal, against a two-sided alternative hypothesis. Give appropriate hypotheses (in words and symbols), test statistic, p-value, and conclusion.

39. A person suspects that a certain grocery store is overstating the weight of fresh chickens, and are thus overcharging their customers. He randomly samples five chicken packages, records the weight stated on the package (in grams) and the true weight of the chicken. The results are illustrated in Table 10.10.

Even though it is somewhat dubious to use the t procedures for such a small sample size, the person conducting the investigation feels the normality assumption is reasonable, and goes ahead with the t procedures. He find the output:

Chicken	Stated weight	Actual weight
1	908	866
2	1120	1087
3	795	783
4	912	890
5	1402	1397

Table 10.10: Stated weight and actual weight for a sample of five chickens.

```
        Paired t-test
data:  stated and actual
t = 3.3861, df = 4, p-value = 0.02763
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  4.104811 41.495189
sample estimates:
mean of the differences
                   22.8
```

(a) Give a summary of the results of the analysis.

(b) If the normality assumption is not in fact reasonable, what are the consequences?

40. Researchers are investigating whether a new experimental method of measuring viscosity in gases is as effective as an older method. They take measurements of viscosity on 6 samples of gas, using both the new method and the older method. The results are given in Table 10.11. The researchers want to test whether there is a difference between the two methods, and estimate the size of the difference with a confidence interval.

Gas Sample	Old Method	New Method
1	2.74	2.71
2	3.11	3.01
3	2.99	3.11
4	2.99	2.94
5	2.88	2.96
6	3.29	3.21

Table 10.11: Measures of viscosity in gases.

(a) Explain why a paired-difference procedure should be used in this situation.

(b) If we want to use a t inference procedure, how could we assess if the normality assumption is reasonable?

For the following questions, suppose that the normality assumption is reason-

able.

(c) Calculate a 95% confidence interval for the population mean difference, and give a proper interpretation of the interval.

(d) Conduct the appropriate hypothesis test. Give appropriate hypotheses (in words and symbols), test statistic, p-value, and conclusion.

41. The reaction times of 4 students are measured before and after having several alcoholic beverages. The results are shown in Table 10.12. Suppose we wish to test the null hypothesis that alcohol has no effect on reaction times, against the alternative that it slows reaction times. Assume that the differences in the times are normally distributed, and use the appropriate t procedure. (But keep in mind that the use of a t procedure is questionable for such a small sample size.)

Student	Before	After
1	0.41	0.78
2	0.38	0.63
3	0.52	0.94
4	0.41	0.47

Table 10.12: Before and after reaction times.

(a) Carry out the hypothesis test. Give appropriate hypotheses, p-value, test statistic and conclusion at the 5% significance level.

(b) Calculate a 95% confidence interval for the mean change in reaction times. Give a proper interpretation of the interval.

(c) What are the advantages to setting up an experiment this way, instead of splitting up the students into 2 independent groups?

(d) What are the disadvantages to setting up an experiment in this way?

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. What are the mean and variance of the sampling distribution of $\hat{p}$?

2. For what values of n and p does the normal approximation to the distribution of $\hat{p}$ work best? For what values of n and p is the normal approximation very poor?

11.3 Confidence Intervals and Hypothesis Tests for the Population Proportion p

3. When do we create a confidence interval for $\hat{p}$?

4. A random sample of 200 observations from a population revealed that 62 individuals had a certain characteristic.

   (a) What is $\hat{p}$?
   (b) What is $SE(\hat{p})$?
   (c) What is a 95% confidence interval for p?
   (d) If we wish to test the null hypothesis that $p = 0.5$, what is $SE_0(\hat{p})$?

(e) What is the value of the appropriate test statistic?

5. In each of the following scenarios, state if it is reasonable to use the normal approximation to calculate a confidence interval for p. (Use the $n\hat{p} \geq 15, n(1-\hat{p}) \geq 15$ guideline.)

   (a) $\hat{p} = 0.70$, $n = 10$.
   (b) $\hat{p} = 0.70$, $n = 200$.
   (c) $\hat{p} = 0.99$, $n = 200$.
   (d) $\hat{p} = 0.01$, $n = 10{,}000$.
   (e) $\hat{p} = 0$, $n = 200$.

6. A study[1] of births in Liverpool, UK, investigated a possible relationship between parental smoking status during pregnancy and the likelihood of a male birth. In one part of the study, researchers drew a sample of 363 births in which both parents were heavy smokers during the pregnancy. Of the 363 babies born to these couples, 158 were male. Suppose we wish to construct a 95% confidence interval for the population proportion of male births to heavy-smoking parents in Liverpool.

   (a) What does p represent in this case?
   (b) What is the point estimate of p?
   (c) Is it reasonable to use large sample methods to calculate a confidence interval for p here?
   (d) What is the standard error of the sample proportion?
   (e) What is a 95% confidence interval for p?
   (f) Give an interpretation of the 95% confidence interval for p.
   (g) Test the null hypothesis that the population proportion of male births to heavy-smoking parents is 0.50, against a two sided alternative. Give appropriate hypotheses, standard error, value of the test statistic, p-value, and conclusion at $\alpha = 0.05$.
   (h) To what population do your conclusions apply?

11.4 Determining the Minimum Sample Size n

7. Find the minimum sample size required in each of the following situations.

   (a) We wish to estimate p within 0.03 with 95% confidence, and we have no reasonable estimate of p beforehand.

[1] Koshy et al. (2010). Parental smoking and increased likelihood of female births. *Annals of Human Biology*, 37(6):789–800.

(b) We wish to estimate p within 0.03 with 95% confidence, and from prior information we feel strongly that p is approximately 0.20.
(c) We wish to estimate p within 0.01 with 90% confidence, and we have no reasonable estimate of p beforehand.
(d) We wish to estimate p within 0.01 with 99% confidence, and we have no reasonable estimate of p beforehand.
(e) We wish to estimate p within 0.01 with 99% confidence, and we know for certain that p lies between 0.1 and 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. In words, what is the meaning of $SE(\hat{p}_1 - \hat{p}_2)$?

9. A random sample of 400 observations from population 1 revealed that 82 individuals had a certain characteristic. A random sample of 400 observations from population 2 revealed that 104 individuals had that characteristic.

   (a) What are the values of $\hat{p}_1$ and $\hat{p}_2$?
   (b) What is $SE(\hat{p}_1 - \hat{p}_2)$?
   (c) Calculate a 95% confidence interval for $p_1 - p_2$.
   (d) Test the null hypothesis that the population proportions are equal, against a two-sided alternative hypothesis. Give appropriate hypotheses, value of the pooled proportion $\hat{p}$, value of the standard error, test statistic and p-value. Is there significant evidence against the null hypothesis at $\alpha = 0.05$?

10. Much research has gone into studying how homing pigeons are able to navigate to their home loft from an unfamiliar release point, but the precise mechanisms are still unknown. It is known that homing pigeons can detect the earth's magnetic field, and that is likely a contributing factor in their ability to navigate. A study[2] investigated this in an experiment involving 77 homing pigeons. The pigeons were randomly divided into a magnetic pulse group (group M) and a control group (group C). The 38 control group pigeons were released at a location 106 km from the home loft, and 22 found their way home. The 39 members of the magnetic pulse group

[2] Holland et al. (2013). A magnetic pulse does not affect homing pigeon navigation: a gps tracking experiment. *The Journal of Experimental Biology*, 216:2192–2200.

received a strong magnetic pulse (perpendicular to the earth's magnetic field) before being released from the same location. Twenty-one of the 39 magnetic pulse group pigeons made it back to the home loft.

(a) Calculate a 95% confidence interval for the difference between the population proportion of pigeons that navigate home after being subjected to the magnetic pulse, and the population proportion for the control group. (Hint to ease the calculation burden: $SE(\hat{p}_M - \hat{p}_C) = 0.1131$.)
(b) Give an appropriate interpretation of the interval found in 10a.
(c) Perform a hypothesis test of the null hypothesis that the two groups have the same likelihood of making it back to their home loft. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion. (One could make an argument for the one-sided alternative hypothesis that the magnetic pulse reduces the probability of a pigeon arriving home, but play it safe and use a two-sided alternative hypothesis.) (Hint to ease the calculation burden: $SE_0(\hat{p}_M - \hat{p}_C) = 0.1132$.)

11.6 Chapter Exercises

11.6.1 Basic Calculations

11. Calculate the p-value in the following situations.

(a) H_0: $p = 0.3$, H_a: $p > 0.3$, $Z = -1.40$.
(b) H_0: $p = 0.3$, H_a: $p < 0.3$, $Z = -1.40$.
(c) H_0: $p = 0.3$, H_a: $p \neq 0.3$, $Z = -1.40$.
(d) H_0: $p = 0.6$, H_a: $p < 0.6$, $Z = -1.88$.
(e) H_0: $p = 0.6$, H_a: $p \neq 0.6$, $Z = -1.88$.

11.6.2 Concepts

12. What is the difference in meaning of the symbols $\hat{p}$ and p?

13. Would it ever make sense to test H_0: $\hat{p} = 0.25$?

14. What is the meaning of the term $SE(\hat{p})$? Does the term $SE(p)$ have a similar meaning?

15. In each of the following scenarios, state whether using the normal approximation to carry out a hypothesis test is reasonable. (Use the $np_0 \geq 15, n(1 - p_0) \geq 15$

guideline.)

(a) H_0: $p = 0.20$, $n = 10$.
(b) H_0: $p = 0.20$, $n = 200$.
(c) H_0: $p = 0.9980$, $n = 10$.
(d) H_0: $p = 0.9980$, $n = 1{,}000$.
(e) H_0: $p = 0.9980$, $n = 100{,}000$.

16. An Ipsos-Reid poll asked Canadians whether schools in their community have become "less safe" than they were five years ago. Fifty percent of the respondents said schools were less safe than 5 years ago. The corresponding confidence interval for p was found to be $(0.46, 0.54)$.

    There are some potential sources of bias in surveys like this one. For example, the question could be biased (e.g. "With the increase in violent behaviour, schools have become less safe, wouldn't you agree?") The folks at Ipsos are professionals, and wouldn't give questions this blatantly biased. But if our hope is to estimate the proportion of *all* Canadians who *feel* that schools have becomes less safe, what are some other potential sources of bias?

17. 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) $\hat{p}$ is an unbiased estimator of p.
    (b) When $n < 30$ we should use the t distribution when calculating confidence intervals for p.
    (c) The true distribution of $\hat{p}$ is based on the binomial distribution.
    (d) The true standard deviation of the sampling distribution of $\hat{p}$ depends on the value of p.
    (e) The sampling distribution of $\hat{p}$ is perfectly normal for large sample sizes.

18. 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 normal approximation to the sampling distribution of $\hat{p}$ works best when we have a large sample size and $p = 0.5$.
    (b) The sampling distribution of $\hat{p}$ becomes more normal as p tends to 1.
    (c) The sampling distribution of p is approximately normal for large sample sizes.
    (d) All else being equal, the value of $SE(\hat{p})$ decreases as the sample size increases.
    (e) In repeated sampling, exactly 95% of 95% confidence intervals for p will capture p.

11.6.3 Applications

19. Consider again the information in Question 10. In part of the experiment, 38 control group pigeons were released from an unfamiliar site 106 km from their home loft, and 22 of these pigeons were able to successfully navigate home. Suppose we wish to construct a confidence interval for the true proportion of control group pigeons that will find their way to their home loft (under the conditions of the study).

    (a) What does p represent in this case?
    (b) What is the point estimate of p?
    (c) Is it reasonable to use large sample methods to calculate a confidence interval for p here?
    (d) What is the standard error of the sample proportion?
    (e) What is a 95% confidence interval for p?
    (f) Give a proper interpretation of the 95% confidence interval.
    (g) Suppose that, before the study was conducted, a scientist claimed that the true proportion of control group pigeons that would navigate back to their home loft would be no more than 0.25. Does this study yield strong evidence against this scientist's claim? Test the null hypothesis that the population proportion is 0.25, against the appropriate alternative hypothesis. Give appropriate hypotheses, standard error, value of the test statistic, p-value, and conclusion.

20. Seat belt marks (bruising) on the body are sometimes used as evidence that a person was wearing a seat belt during a car crash. A study[3] investigated seat belt marks in victims of fatal car crashes in Sydney, Australia. (The authors looked only at cases where there was no airbag deployment.) In 74 fatalities in which the victim was wearing a seat belt, 27 victims showed seat belt marks.

    (a) What does p represent in this case?
    (b) What is the point estimate of p?
    (c) Is it reasonable to use large sample methods to calculate a confidence interval for p here?
    (d) What is the standard error of the sample proportion?
    (e) What is a 95% confidence interval for p?
    (f) Give an appropriate interpretation of the confidence interval for p.
    (g) Is there a hypothesis that should be tested here?

21. A study[4] investigated various characteristics of Finnish murders. In one part of the

[3] Chase et al. (2007). Safety restraint injuries in fatal motor vehicle collisions. *Forensic Science, Medicine, and Pathology*, 3:258–263.

[4] Häkkänen et al. (2009). Gender differences in Finnish homicide offence characteristics. *Forensic Science International*, 186:75–80

study, a random sample of 91 Finnish male convicted murderers was drawn (there was a single murder in each case). Twenty-four of these 91 murders were of the offender's domestic partner. Suppose we wish to estimate the true proportion of murders that are of the domestic partner of the offender (for murders committed by males).

(a) What does p represent in this case?
(b) What is the point estimate of p?
(c) Is it reasonable to use large sample methods to calculate a confidence interval for p here?
(d) What is the standard error of the sample proportion?
(e) What is a 95% confidence interval for p?
(f) Give an interpretation of the 95% confidence interval for p.
(g) Suppose it was previously believed that in 40% of murders committed by a Finnish male, the victim was the offender's domestic partner. Test the null hypothesis that the population proportion is 0.40, against a two sided alternative. Give appropriate hypotheses, standard error, value of the test statistic, p-value, and conclusion at $\alpha = 0.05$.
(h) To what population do the conclusions apply?

22. In another aspect of the study of Finnish murders first discussed in Question 21, researchers drew random samples of murders committed by males and murders committed by females. Of 91 murders committed by a female, in 32 cases the victim was their domestic partner. Of 91 murders committed by a male, in 24 cases the victim was their domestic partner.

(a) Calculate a 90% confidence interval for the difference in the population proportions between female and male murderers. (Hint to ease the calculation burden: $SE(\hat{p}_F - \hat{p}_M) = 0.0681$.)
(b) Give an appropriate interpretation of the interval found in 22a.
(c) Perform a hypothesis test of the null hypothesis that, in Finnish murders, the proportion of victims that are the domestic partner of the offender is the same for male and female offenders, against the alternative that these proportions are different. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion at the 5% level of significance. (Hint to ease the calculation burden: $SE_0(\hat{p}_F - \hat{p}_M) = 0.0684$.)

23. A study[5] at a medical clinic in Iran investigated a possible association between peptic ulcers and a variety of factors. The study involved a sample of 60 peptic ulcer sufferers from a medical clinic, and a sample of 44 apparently healthy volunteers to serve as the control group. In one part of the study, the authors investigated

[5]Mahdavi et al. (2013). Is there any relationship between leb antigen expression and *Helicobacter pylori* infection? *Blood Cells, Molecules and Diseases*, 51:174–176.

a possible difference between the groups in their Lewis B (Le^b) blood group antigen expression. (Previous studies had shown a possible relationship between Le^b expression and a mechanism for peptic ulcers.) Forty-three of the 60 peptic ulcer sufferers had Le^b expression. Twenty-seven of the 44 control group members had Le^b expression.

(a) Calculate a 95% confidence interval for the difference in the population proportions between peptic ulcer patients and the control group. (Hint to ease the calculation burden: $SE(\hat{p}_U - \hat{p}_C) = 0.0937$.)
(b) Give an appropriate interpretation of the interval found in 23a.
(c) Perform a hypothesis test of the null hypothesis that the proportion of individuals with Le^b expression is the same for the two groups, against the alternative hypothesis that the proportions are different. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion at the 5% level of significance. (Hint to ease the calculation burden: $SE_0(\hat{p}_U - \hat{p}_C) = 0.0931$.)

24. Consider again the study first discussed in Question 6, which investigated a possible relationship between parental smoking status during pregnancy and the likelihood of a male birth. Of 363 male births in which both parents were heavy smokers during the pregnancy, 158 of the babies were male. Of the 5045 births in which both parents were non-smokers during the pregnancy, 2685 of the babies were male.

(a) Calculate a 95% confidence interval for the difference between the population proportion of male births for heavy-smoking parents, and the population proportion of male births for non-smoking parents. (Hint to ease the calculation burden: $SE(\hat{p}_H - \hat{p}_N) = 0.0270$.)
(b) Give an appropriate interpretation of the interval found in 24a.
(c) Perform a hypothesis test of the null hypothesis that the male birth rate is the same for both heavy-smoking and non-smoking parents. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion at the 5% level of significance. (Hint to ease the calculation burden: $SE_0(\hat{p}_H - \hat{p}_N) = 0.0271$.)
(d) To what population do your conclusions apply? Do the results of this study imply strong evidence that parents smoking *cause* a decrease in the proportion of male births?

25. Does a pleasant ambient fragrance increase the success rate of a courtship request? A study[6] investigated this question by having a good-looking 20 year-old man approach 18-25 year-old women walking alone in a mall. On 200 occasions, the man approached as the woman was walking in an area of "pleasant ambient odours (e.g., pastries)", and on 200 occasions he approached in an area of no odour. In every case, the man approached the woman, made a brief scripted statement including a

[6] Guéguen et al. (2012). The sweet smell of. courtship: Effects of pleasant ambient fragrance on womenÕs receptivity to a manÕs courtship request. *Journal of Environmental Psychology*, 32:123–125

compliment, and then asked for the woman's phone number. Of the 200 cases in areas of pleasant ambient odour, 46 women gave their phone number. Of the 200 cases in areas of no ambient odour, 27 women gave their phone number.

(a) Calculate a 95% confidence interval for the difference between the success rate in areas with a pleasant fragrance and the success rate in areas with no odour $(p_F - p_C)$. (Hint to ease the calculation burden: $SE(\hat{p}_F - \hat{p}_C) = 0.0383$.)

(b) Give an appropriate interpretation of the interval found in 25a.

(c) Perform a hypothesis test of the null hypothesis that the likelihood of getting a phone number is the same in both areas. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion at the 5% level of significance. (Hint to ease the calculation burden: $SE_0(\hat{p}_F - \hat{p}_C) = 0.0386$.)

26. Part of the Salk polio vaccine trials in 1954 involved a massive double blind randomized experiment, in which approximately 400,000 elementary school children were randomly assigned to either the Salk polio vaccine or a placebo group. The results:

	Placebo	Vaccine
Polio	160	86
No polio	200270	199661
Total	200430	199747

(The counts have been slightly modified from the original data.)

(a) Let p_p represent the population proportion of those in the placebo group that would develop polio under these conditions, and p_v represent the population proportion of those in the vaccine group that would develop polio under these conditions. What is the point estimate of the difference in the population proportions? (What is the point estimate of $p_p - p_v$?)

(b) If we wish to test the null hypothesis that the population proportions are equal, what is the value of the appropriate Z test statistic? (Hint to ease the calculation burden: $SE_0(\hat{p}_p - \hat{p}_v) = 0.00007836327$.)

(c) What is the p-value of the test of the null hypothesis of equal population proportions against a two-sided alternative hypothesis?

(d) Give a summary of the results of the hypothesis test.

(e) Since polio is a rare disease, the *difference* in population proportions may not be the best way of comparing the two proportions. In many situations, especially with rare events, the *relative risk* is the more informative quantity. In this case the relative risk would be the ratio of the estimated probabilities of developing polio. That is, $RR = \frac{\hat{p}_{placebo}}{\hat{p}_{vaccine}}$. If we have a relative risk of 3, say,

then that means individuals given the placebo would be 3 times more likely to develop polio. In many cases this is a better and more intuitive measure of the difference between proportions. What is the relative risk of developing polio?

11.6.4 Extra Practice Questions

27. The median lethal dose (often called LD50 (lethal dose, 50%) or LC50 (lethal concentration, 50%)), is often reported in toxicology studies. It is the dose that results in 50% mortality within a specified time period. Suppose you are investigating the effect of zinc concentration on the mortality of a certain type of freshwater minnow. As part of this study, you subject 100 of these minnows to a zinc concentration of 5 ppm. Within 48 hours, 68 of the minnows are dead.

    (a) Construct a 95% confidence interval for the true proportion of this type of minnow that would be dead within 48 hours if exposed to 5 ppm of zinc.

    (b) Based on this interval, is there a strong indication that the 48 hour LD50 is different from 5 ppm?

    (c) Carry out a hypothesis test of the null hypothesis that 5 ppm of zinc will kill 50% of this type of minnow within 48 hours, against a two-sided alternative hypothesis. Give appropriate hypotheses, standard error, test statistic, p-value, and conclusion at $\alpha = 0.05$.

    (d) Suppose in a different experiment, with a different type of minnow, we wished to estimate the proportion that would be dead in 48 hours to within 0.03, with 99% confidence. What is the minimum sample size that would be required? (Suppose that we have no reasonable estimate of the true proportion before the study, and thus we must use the conservative estimate of p in the formula.)

    (e) Suppose that in another part of this study, after exposure to a high level of zinc, 198 of 200 minnows are dead within 48 hours. Would it be reasonable to use the normal approximation to construct a confidence interval for p in this scenario?

28. Suppose it was previously believed that in a certain geographical area, approximately 50% of babies born prematurely at 25 weeks did not survive until at least their first birthday. You investigate this claim and find data on 67 babies born at 25 weeks. 43 of these babies survived until they were one year old. Suppose it is reasonable to assume that your data represents a random sample of births from that area at that time.

(a) Construct a 95% confidence interval for the proportion of babies born at 25 weeks that survive until they are at least one year old. Give an appropriate interpretation of the interval.

(b) Test the null hypothesis that half of babies born at 25 weeks die, against a two-sided alternative hypothesis. Give appropriate hypotheses, test statistic, p-value, and conclusion.

29. A random sample of 40 soft-serve ice cream vendors in a large city revealed that 17 of them had *E. coli* counts in excess of recommended guidelines. Calculate a 95% confidence interval for the population proportion, and give a proper interpretation of the interval.

30. The 2006 Canadian census revealed that approximately 23% of Canadians between 25 and 64 years of age have a university degree. In a certain area at that time, a random sample of 500 adults in this age group revealed that 25 had a university degree.

    (a) Construct a 90% confidence interval for the proportion of people between 25 and 64 in this area at that time that had a university degree.

    (b) Test whether the proportion of people between 25 and 64 in this area with a university degree differed from the rest of Canada. Give appropriate hypotheses, test statistic, p-value, and conclusion.

31. You are interested in using a marketing campaign to sell new cars to recent university graduates. Before starting the campaign, you would like to have some idea of the proportion of graduating students who intend to buy a new car in the next year. If you wish to estimate this proportion to within 0.04 with 90% confidence, what is the minimum sample size that is required? Assume that you do not have a good estimate of the proportion before taking the sample, and thus you need to use the most conservative estimate of the sample size required.

32. Pollsters were interested in possible differences between men and women in their presidential approval ratings. 100 men and 100 women were asked if they approved of the way the president was handling his job. 58 of the women and 43 of the men said they approved.

    (a) Calculate a 90% confidence interval for the difference in the population proportions between females and males. (Hint to ease the calculation burden: $SE(\hat{p}_F - \hat{p}_M) = 0.0699$.)

    (b) Give an appropriate interpretation of the interval found in 32a.

    (c) Perform a hypothesis test of the null hypothesis that the approval ratings for men and women are the same, against the alternative that they are different. Give appropriate hypotheses, value of the test statistic, p-value, and a conclusion at the 5% level of significance. (Hint to ease the calculation burden:

$SE_0(\hat{p}_F - \hat{p}_M) = 0.070707.$)

33. A company is testing a new heat treatment intended to eradicate bedbugs. As part of the investigation process, they wish to assess what effect different temperatures have on killing bedbugs. They subject 200 adult bedbugs to a temperature of 45 degrees Celsius for 30 minutes, and find that 176 died. They subject 200 different adult bedbugs to a temperature of 47 degrees Celsius for 30 minutes, and find that 184 died.

    (a) What is a 95% confidence interval for the difference in population proportions? Also give an appropriate interpretation of the interval. (Hint to ease the calculation burden: $SE(\hat{p}_1 - \hat{p}_2) = 0.02993.$)

    (b) Suppose we wish to test the null hypothesis that the population proportions are equal, against the alternative that the higher temperature kills a higher proportion of bedbugs. Give appropriate hypotheses in words and symbols, standard error, value of the test statistic, and conclusion at a $\alpha = 0.01$ significance level. (Hint to ease the calculation burden: $SE_0(\hat{p}_1 - \hat{p}_2) = 0.030.$)

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. Suppose we have a sample of 400 observations that are categorized into 4 groups. The first group contains 52 of the observations, the second 149, the third 99, and the fourth 100 observations. Suppose that we wish to test the null hypothesis that these four groups are equally likely.

   (a) What are the appropriate hypotheses, in both words and symbols?
   (b) What are the observed counts? What are the expected counts under the null hypothesis?
   (c) What is the value of the test statistic?
   (d) What are the appropriate degrees of freedom?
   (e) What is the p-value of the test?
   (f) Can the null hypothesis be rejected at $\alpha = 0.10$? At $\alpha = 0.01$?

2. Does the distribution of birth dates of Canadian NHL players differ from that of Canadians as a whole? In his popular book *Outliers*, Malcolm Gladwell made the claim that Canadian children born earlier in the year (January, February, March)

have a better chance of making the NHL when compared to children born later in the year. Does NHL data from 2010/2011 support this claim? Table 13.1 illustrates the distribution of a sample of 510 Canadian NHL player birthdays and the distribution for all Canadians.[12]

Birth date	Q1	Q2	Q3	Q4	Total
Count (NHL)	135	146	124	105	510
Percentage (NHL)	26.5	28.6	24.3	20.6	100
Percentage (Canadians)	24.2	26.0	26.0	23.8	100

Table 13.1: Birthdays of Canadian NHL players and Canadians by quarter.

(a) Is the observed percentage of Q1 birthdays higher for Canadian NHL players than for Canadians as a whole? Is the observed percentage of Q4 birthdays lower for Canadian NHL players than for Canadians as a whole?
(b) Suppose we wish use a χ^2 test to test the hypothesis that the distribution of birthdays for Canadian NHL players is the same as that of Canadians as a whole. What are the appropriate hypotheses of the test?
(c) If H_0 were true, what would be the expected count of NHL player birthdays in Q1? Is this higher or lower than the observed count?
(d) What are the degrees of freedom for the χ^2 test statistic?
(e) Calculate the value of the χ^2 test statistic.
(f) What is the p-value of the test?
(g) Give an appropriate conclusion to the hypothesis test at the 5% significance level.

13.2.2 Testing Goodness-of-Fit for Specific Parametric Distributions

3. Suppose that a type of bird lays exactly three eggs per year (this is unrealistic of course, but play along). A scientist claims that the number of eggs that hatch per nest follows a binomial distribution with $n = 3$, $p = 0.80$. You do not believe this scientist, and follow a sample of 500 nests until the eggs hatch. The results for the 500 nests are summarized in the Table 13.2.

   Does this sample provide strong evidence against the scientist's claim that the number of eggs that hatch in a nest has a binomial distribution with $n = 3$, $p =$

[1]The NHL birth date data was pulled on March 24 2011 using a search on www.nhl.com for current NHL players with Canadian birthplaces.

[2]Q1 is defined to be January, February March, Q2 is defined to be April, May, June, Q3 is defined to be July, August, September, Q4 is defined to be October, November, December. The distribution of birth dates for all Canadians was based on information from Statistics Canada.

Number that hatch	0	1	2	3
Frequency	80	29	164	227

Table 13.2: Number of eggs that hatch per nest.

0.80? Test the claim by performing a chi-square goodness-of-fit test. (Hint: To get the expected counts under the null hypothesis, use the binomial formula with the hypothesized values ($n = 3$, $p = 0.80$), and multiply the probabilities by 500. Note that sometimes we use the data to estimate p, but here the value of p is specified in the null hypothesis.)

(a) What are the appropriate hypotheses?
(b) What are the expected counts?
(c) What is the value of the appropriate test statistic?
(d) What are the appropriate degrees of freedom?
(e) What is the p-value of the test?
(f) Give an appropriate conclusion at the 5% significance level.

4. Consider again the information in the Question 3. Suppose a different scientist has claimed that the number of eggs that hatch should follow a binomial distribution (with an unknown value of p). Does the sample data given in the previous question provide strong evidence against the scientist's claim that the number of eggs that hatch per nest follows a binomial distribution? Test the claim by performing a χ^2 goodness-of-fit test. (The difference between this question and the previous one is that here we must use the data to estimate p, which is the more likely scenario in the real world.)

5. Consider again the information in the Question 3. Suppose a different scientist has suggested that the 4 possible outcomes (0 hatch, 1 hatch, 2 hatch, 3 hatch) are all equally likely. Using the data give in Question 3, test the null hypothesis that the 4 possible outcomes are all equally likely. Give appropriate hypotheses, test statistic, p-value, and conclusion.

13.3 χ^2 Tests for Two-Way Tables

6. Suppose we have a random sample of 1000 observations, categorized according to two variables. The results are given in Table 13.3.

(a) What are the appropriate hypotheses?
(b) What are the expected counts in each cell?
(c) What is the value of the test statistic?

	C1	C2
R1	50	150
R2	200	20
R3	400	180

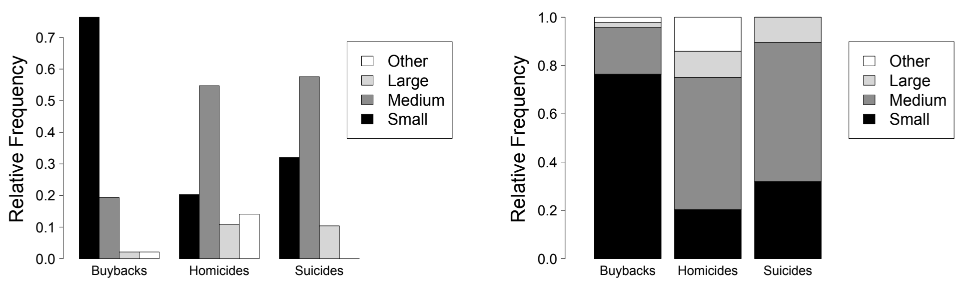
Table 13.3: A 3×2 contingency table.

(d) What are the appropriate degrees of freedom?
(e) What is the p-value of the test?
(f) Is the null hypothesis rejected at $\alpha = 0.10$? At $\alpha = 0.01$?

7. Many cities in the United States have buyback programs for handguns, in which the police department pays people to turn in guns. The guns are then destroyed. Is there a difference between the distribution of the size of guns turned in during buyback programs and the distribution of the size of guns used in homicides and suicides? A study[3] investigated this question, using data from a gun buyback program and police records in Milwaukee. Table 13.4 and Figure 13.1 illustrate the distribution of the calibre of the gun (small, medium, large, other) in the different scenarios.

Gun Calibre	Buybacks	Homicides	Suicides	Total
Small	719	75	40	834
Medium	182	202	72	456
Large	20	40	13	73
Other	20	52	0	72
Total	941	369	125	1435

Table 13.4: Gun calibre for buyback guns and guns used in homicides and suicides.



(a) Side-by-side barcharts. (b) Stacked barcharts.

Figure 13.1: The distribution of calibre of gun for buybacks, homicides, and suicides.

[3]Kuhn, E., Nie, C., O'Brien, M., Withers, R. L., Wintemute, G., and Hargarten, S. W. (2002). Missing the target: a comparison of buyback and fatality related guns. *Injury Prevention*, 8:143–146

(a) What percentage of buyback guns were of small calibre? What percentage of homicide guns were of small calibre?
(b) Suppose we wish to investigate a possible association between gun calibre and gun type (buyback, homicide, or suicide) using a χ^2 test. What are the appropriate hypotheses of the test?
(c) If H_0 were true, what would be the expected count of small calibre buyback guns?
(d) What are the degrees of freedom for the χ^2 test statistic?
(e) The value of the χ^2 test statistic is 422.48. (You should be able to calculate this value, but the calculations are rather long and tedious.) What is the p-value of the test?
(f) Give an appropriate conclusion to the hypothesis test.
(g) Looking at Table 13.4 and Figure 13.1, what are the major differences between the distribution of calibre of gun between the buyback guns and the guns used in homicides or suicides?
(h) To what population do your conclusions apply?

13.4 Chapter Exercises

13.4.1 Basic Calculations

13.4.2 Concepts

8. What will happen if we carry out a χ^2 test and one or more of our expected counts is very small (less than one, say)?

(a) The world will end.
(b) We cannot calculate the test statitistic.
(c) The test statistic will be negative.
(d) Our conclusion will be the exact opposite of what it should be.
(e) The χ^2 approximation may not be very good, possibly resulting in misleading conclusions.

9. Fifty children were asked to taste 5 difference ice creams, and were asked which flavours they liked. The results:

Vanilla	Chocolate	Cherry & Mango	Double Mango	"All Natural Mango" (with simulated mango flavour)
32	42	16	15	14

Would it be reasonable to use our usual χ^2 test here? Why or why not?

10. (a) If we let X represent a random variable that has a χ^2 distribution with 8 degrees of freedom, what is $P(X > 21.955)$?
    (b) If we let X represent a random variable that has a χ^2 distribution with 2 degrees of freedom, what is $P(X > 3.794)$?
    (c) If Z_1, Z_2, Z_3, Z_4 are independent standard normal random variables, what is $P(Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 > 7.779)$?
    (d) If Z_1, Z_2, Z_3 are independent standard normal random variables, what is $P(Z_1^2 + Z_2^2 + Z_3^2 < 11.345)$?

11. 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 χ^2 tests for count data, all else being equal, the greater the value of the test statistic the smaller the p-value.
    (b) The χ^2 distribution tends toward the standard normal distribution as the degrees of freedom increase.
    (c) The χ^2 distribution is skewed to the right.
    (d) In χ^2 tests for count data, we lose 1 degree of freedom for each parameter we estimate using our data.
    (e) χ^2 statistics can take on negative values.

12. 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 χ^2 test statistic has exactly a χ^2 distribution.
    (b) The mean of a χ^2 distribution is equal to its degrees of freedom.
    (c) For a given set of expected counts in a two-way table of a given size, the greater the difference between the observed and expected counts, the greater the value of the χ^2 test statistic.
    (d) For a given set of expected counts in a two-way table of a given size, the greater the difference between the observed and expected counts, the smaller the p-value.
    (e) For a given set of expected counts in a two-way table of a given size, the greater the difference between the observed and expected counts, the greater the evidence against the null hypothesis.
    (f) For a 2×2 table, the value of the χ^2 test statistic is the square of the corresponding Z test statistic.

13.4.3 Applications

13. In a study[4] investigating genetic linkage in tomato plants, the researchers crossed one pure line of tomato plants that were tall and cut-leaf with another pure line that were dwarf and potato-leaf. Tall is genetically dominant to dwarf, and cut-leaf is genetically dominant to potato-leaf, so the tomato plants resulting from this first generation cross were all tall and cut-leaf (they had the tall/cut-leaf *phenotype*). This first generation was then self-crossed. For the second generation, under Mendelian inheritance with independent assortment, a 9:3:3:1 ratio of tall/cut-leaf:tall/potato-leaf:dwarf/cut-leaf:dwarf/potato-leaf phenotypes would be expected. A ratio very different from this would indicate that the gene controlling the height and the gene controlling the leaf type are not being inherited independently. The counts of observed phenotypes for the 1611 tomato plants are given in Table 13.5.

	Phenotype			
	Tall/cut-leaf	Tall/potato-leaf	Dwarf/cut-leaf	Dwarf/potato-leaf
Observed count	926	288	293	104

Table 13.5: Phenotypes observed in the second generation.

Suppose we wish to test the null hypothesis that these 4 phenotypes occur in a 9:3:3:1 ratio.

(a) What are the appropriate hypotheses?
(b) What is the value of the χ^2 test statistic?
(c) What is the p-value of the test?
(d) Give an appropriate conclusion to the test.

14. In genetics experiments involving pea plants (like the famous ones conducted by Mendel), one variable that is often measured is the shape of the ripe seeds (the seeds are either round or wrinkled). Round seeds (R) are dominant over wrinkled seeds (r) so a pea plant will have round seeds if it is homozygous for round seeds (RR) or heterozygous (Rr), and will have wrinkled seeds only if it is homozygous for wrinkled seeds (rr). If heterozygous plants are self-crossed, then on average 75% of the resulting plants will have round seeds and 25% will have wrinkled seeds.

Suppose you suspect that a field of pea plants is heterozygous for seed shape (Rr), but you are not sure. (Perhaps you were conducting an experiment and lost your notes about what was on that field.) These plants are self-crossed, and of the resulting 80 plants, 72 have round seeds and 8 have wrinkled seeds. Does this give strong evidence that the true proportion of plants with round seeds from this crossing differs from 0.75? You may wish to use a χ^2 test to investigate this question.

[4]MacArthur, J. (1931). Linkage studies with the tomato. *Transactions of the Royal Canadian Institute*, 18:1–19.

(a) What are the appropriate hypotheses?
(b) What is the value of the χ^2 test statistic?
(c) What is the p-value of the test?
(d) What is an appropriate conclusion at the 5% significance level?

15. Consider again the information in Question 14. In this question the null hypothesis that the true proportion of plants with round seeds is 0.75 was tested with a χ^2 test. This null hypothesis could also be tested with a Z test for a single proportion.

(a) What is the value of the Z test statistic?
(b) What is the p-value of the Z test for a two-sided alternative hypothesis?
(c) What is the relationship between the Z statistic and the χ^2 test statistic found in 14b?
(d) What is the relationship between the p-value of the Z test and the p-value of the χ^2 test found in part 14c?

16. Benford's law describes the distribution of the first digit in the value of a variable. (The first digit refers to the first non-zero digit in the number. For example, the first digit of 312.7 is 3, and the first digit of 0.072 is 7.) In many naturally occurring scenarios, the digits 1–9 are not equally likely to be the first digit (smaller first digit values are more likely). This has been shown for many different variables, including bank account balances, populations of cities around the world, and lengths of rivers. (It does not apply to all types of data. For example, artificially structured numbers like telephone numbers and zip codes do not have this feature.) For variables that satisfy Benford's law, the probability that the first digit is d ($d = 1, \ldots, 9$) is $\log_{10}(1 + \frac{1}{d})$. Under Benford's law the first digit has approximately a 30% chance of being a 1, but only a 5% chance of being a 9.

What is the distribution of the first digit in sample means reported in scientific journal articles? Does the distribution follow Benford's law? Or is it closer to a discrete uniform distribution, or something else entirely? A sample of 126 means from journal articles resulted in the first digits reported in Table 13.6 and Figure 13.2. (The means were independently drawn by your author from a variety of scientific journal articles.)

(a) Suppose we wish to test the null hypothesis that the distribution of the first digit in reported sample means is uniform. What are the appropriate hypotheses?
(b) What are the degrees of freedom?
(c) The value of the χ^2 test statistic is 49.29. (The calculation is straightforward, and you should understand how to arrive at this value, but it is a bit of a hassle to calculate.) What is the p-value of the test?
(d) What is an appropriate conclusion to the hypothesis test?

First digit	1	2	3	4	5	6	7	8	9
Observed count	34	22	17	11	11	13	7	4	7
Observed proportion	0.27	0.17	0.13	0.09	0.09	0.10	0.06	0.03	0.06
Uniform proportion	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9
Uniform expected count	14	14	14	14	14	14	14	14	14
Benford proportion	0.30	0.18	0.12	0.10	0.08	0.07	0.06	0.05	0.05
Benford expected count	37.9	22.2	15.7	12.2	10.0	8.4	7.3	6.4	5.8

Table 13.6: Observed counts and proportions, and the counts and proportions that would be expected under a uniform distribution and under Benford's law for 126 sample means.

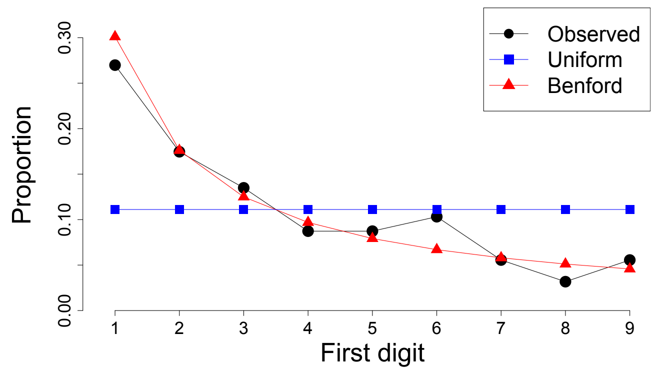


Figure 13.2: The observed proportion, and the expected proportions under a uniform distribution and Benford's law.

(e) The hypothesis test yields extremely strong evidence that the true distribution is not uniform. Use information from Figure 13.2 to describe how the distribution differs from the uniform distribution.

(f) Suppose we wish to test the null hypothesis that the distribution of the first digit in reported sample means follows Benford's law. What are the appropriate hypotheses?

(g) What are the degrees of freedom?

(h) The value of the χ^2 test statistic is 4.41. (The calculation is straightforward, and you should understand how to arrive at this value, but it is a bit of a hassle to calculate.) What is the p-value of the test?

(i) What is an appropriate conclusion?

(j) Are the results of the test consistent with what is illustrated in Figure 13.2?

17. A study[5] involving a random sample of 305 Libyan medical students revealed the following information about their ABO blood type and Rh factor.

[5]Fayrouz et al. (2012). Relation between fingerprints and different blood groups. *Journal of Forensic and Legal Medicine*, 19:18–21.

	A	B	AB	0	Total
Rh +ve	92	30	8	136	266
Rh -ve	9	9	8	13	39
Total	101	39	16	149	305

Table 13.7: Blood types of 305 Libyan medical students.

(a) Suppose we wish to investigate a possible association between ABO blood type and Rh factor using a χ^2 test. What are the appropriate hypotheses of the test?

(b) If H_0 were true, what would be the expected count of individuals with blood type A $-$ve?

(c) What are the degrees of freedom for the χ^2 test statistic?

(d) The value of the χ^2 test statistic is 27.1367. (You should be able to calculate this value, but the calculations are rather long and tedious.) What is the p-value of the test?

(e) Give an appropriate conclusion at the 5% significance level.

(f) To what population do your conclusions apply?

(g) Optional: Why is the conclusion to the test an unusual one? (This requires some background knowledge in blood types.)

18. An Iranian study[6] investigated a possible association between peptic ulcers and a variety of factors. The study involved a sample of 60 peptic ulcer sufferers from a medical clinic, and a sample of 44 apparently healthy volunteers to serve as the control group. In one part of the study, the ABO blood type of the individuals was recorded. The results are listed in Table 13.8.

	A	B	AB	O
Peptic ulcer sufferers	11	16	3	30
Healthy individuals	15	5	5	19

Table 13.8: Blood types of 60 patients with peptic ulcers and 44 healthy individuals.

(a) What percentage of the 60 of peptic ulcer patients have blood type A? What percentage of the 44 healthy individuals have blood type A?

(b) Suppose we wish to investigate a possible association between peptic ulcers and ABO blood type using a χ^2 test. What are the appropriate hypotheses of the test?

(c) If H_0 were true, what would be the expected count of peptic ulcer patients with blood type A?

[6] Mahdavi et al. (2013). Is there any relationship between leb antigen expression and *Helicobacter pylori* infection? *Blood Cells, Molecules and Diseases*, 51:174–176.

(d) What are the degrees of freedom for the χ^2 test statistic?
(e) The value of the χ^2 test statistic is 7.0521. (You should be able to calculate this value, but the calculations are rather long and tedious.) What is the p-value of the test?
(f) Give an appropriate conclusion to the hypothesis test at the 5% level of significance.

19. A study[7] of births in Liverpool, UK, investigated a possible relationship between parental smoking status during pregnancy and the likelihood of a male birth. The results for 6853 births in which the father was a non-smoker are given in Table 13.9.

	No smoking	Light smoking	Heavy smoking	Total
Male birth	2685	735	175	3595
Female birth	2360	708	190	3258
Total	5045	1443	365	6853

Table 13.9: Maternal smoking status during pregnancy and sex of the child for 6853 births. (The father was a non-smoker in each case.)

(a) What is the estimated proportion of male births for non-smoking couples? What is the estimated proportion of male births for couples in which the father is a non-smoker, but the mother is a heavy smoker?
(b) Suppose we wish to investigate a possible association between maternal smoking status and sex of the child using a χ^2 test. What are the appropriate hypotheses of the test?
(c) If H_0 were true, what would be the expected count of male births to non-smoking women?
(d) What are the degrees of freedom for the χ^2 test statistic?
(e) The value of the χ^2 test statistic is 5.499. (You should be able to calculate this value, but the calculations are rather long and tedious.) What is the p-value of the test?
(f) Give an appropriate conclusion to the hypothesis test.
(g) To what population do your conclusions apply?

20. In the same study as given in Question 19, the researchers also investigated a possible association between paternal smoking status during pregnancy and the likelihood of a male birth. The results for 6032 births in which the mother was a non-smoker are given in Table 13.10.

(a) What is the estimated proportion of male births for non-smoking couples? What is the estimated proportion of male births for couples in which the

[7] Koshy et al. (2010). Parental smoking and increased likelihood of female births. *Annals of Human Biology*, 37(6):789–800.

	No smoking	Light smoking	Heavy smoking	Total
Male birth	2685	385	122	3192
Female birth	2360	357	123	2840
Total	5045	742	245	6032

Table 13.10: Paternal smoking status during pregnancy and sex of the child for 6032 births. (The mother was a non-smoker in each case.)

mother is a non-smoker, but the father is a heavy smoker?

(b) Suppose we wish to investigate a possible association between paternal smoking status and sex of the child using a χ^2 test. What are the appropriate hypotheses of the test?

(c) If H_0 were true, what would be the expected count of male births to a non-smoking father?

(d) What are the degrees of freedom for the χ^2 test statistic?

(e) The value of the χ^2 test statistic is 1.461. (You should be able to calculate this value, but the calculations are rather long and tedious.) What is the p-value of the test?

(f) Give an appropriate conclusion to the hypothesis test.

(g) To what population do your conclusions apply?

21. A Finnish study[8] looked at a possible association between the gender of convicted murderers and their relationship to the victim. A random selection of 91 female murderers and a random selection of 91 male murderers revealed the information in Table 13.11.

Gender	Male	Female	Total
Acquaintance	61	37	98
Partner	22	32	54
Family Member	4	18	22
Stranger	4	4	8
Total	91	91	182

Table 13.11: Gender of murderer and their relationship to the offender in Finnish murders.

(a) For female murderers, what percentage of murders were of family members? For male murderers, what percentage of murders were of family members?

(b) Suppose we wish to investigate a possible association between the gender of the murderer and the relationship to the victim using a χ^2 test. What are the appropriate hypotheses of the test?

[8] Häkkänen et al. (2009). Gender differences in Finnish homicide offence characteristics. *Forensic Science International*, 186:75–80.

(c) If H_0 were true, what would be the expected count for female murderers of family members?
(d) What are the degrees of freedom for the χ^2 test statistic?
(e) The value of the χ^2 test statistic is 16.639. (You should be able to calculate this value, but the calculations are rather long and tedious.) What is the p-value of the test?
(f) Give an appropriate conclusion to the hypothesis test.
(g) By looking at the values in the table, what appear to be the differences in the distributions of relationship to the victim between male and female murderers?
(h) To what population do your conclusions apply?

22. Consider again the information from the Polio vaccine trials (which was first discussed in the exercises for inference for proportions).

	Placebo	Vaccine
Polio	160	86
No polio	200270	199661
Total	200430	199747

(a) Carry out a χ^2 test, testing the null hypothesis that the vaccine has no effect. What is the value of the test statistic?
(b) When we analyzed this data using a two sample Z test, we found $Z = 4.6927$. What is the relationship between the two-sample Z test statistic value the χ^2 test statistic value found here?

23. In order to make it a little more difficult for test writers to cheat, a professor has two versions of each test. One version has a white cover sheet and one has a yellow cover sheet. The versions are of equal difficulty level, but the questions are slightly different and the order is changed. In the examination room, students choose where to sit and thus choose whether to write the white or yellow version (the tests were placed around the rooms beforehand).

If a student's ID number is not filled out properly on the computer answer sheet, then it must be tracked down and fixed. In one test, far more writers of the yellow exam gave an incorrect ID number, as illustrated in Table 13.12.

	White	Yellow
Incorrect ID	1	16
Correct ID	422	403

Table 13.12: Number of correctly and incorrectly filled out ID numbers on a statistics test. (This is real data from a 2012 course at the University of Guelph.)

Suppose we wish to test whether there is an association between the colour of the

test and whether or not the student correctly filled in their ID number.

(a) What are the appropriate hypotheses?
(b) What is the value of the χ^2 test statistic?
(c) What is the p-value of the test?
(d) Even though the test yields very strong evidence against the null hypothesis, the professor believed it to simply be a fluke and not a real effect. What are some factors that might make them feel this way?

13.4.4 Extra Practice Questions

24. One hundred volunteers take part in a psychology study. The participants are shown 3 photos of an individual, one after the other, and asked which picture is the most flattering. In reality, all 3 pictures are the same. In the study, 24 individuals picked picture #1, 37 picked picture #2, and 39 picked picture #3. Test the hypothesis that all 3 pictures are equally preferred.

    (a) What are the appropriate hypotheses, in words and symbols?
    (b) What is the value of the test statistic?
    (c) What is the p-value of the test?
    (d) Give an appropriate conclusion.

25. Viewers of a newly released movie were asked to give it a rating from 1 to 4 stars. The results from a sample of 400 viewers are illustrated in Table 13.13.

# of stars	1⋆	2⋆	3⋆	4⋆
# of viewers	12	79	209	100

Table 13.13: Ratings of 400 viewers.

Before the sample was drawn, a movie executive predicted that viewers would give the above ratings in a 1:1:2:2 ratio. That is, she believed the proportions would be 1/6, 1/6, 2/6, 2/6. Does the sample above provide strong evidence against the executive's claim? Perform the appropriate chi-square test.

    (a) What are the appropriate hypotheses, in words and symbols?
    (b) What is the value of the test statistic?
    (c) What is the p-value?
    (d) Give an appropriate conclusion.

26. A manufacturer of potato chips wants to test the popularity of 3 new flavours it has developed. As an initial test, the product development branch wants to test the null hypothesis that all 3 brands are equally preferred by consumers. They

administered a taste test to 120 people, in which each person tasted all 3 potato chip flavours and named their favourite. The results are given in Table 13.14.

Flavour	Salty Tobacco	Sweet BBQ	Raspberry Vinaigrette
# who preferred flavour	26	67	27

Table 13.14: Favourite chip flavours in a taste test.

Suppose we wish to test the null hypothesis that all 3 flavours are equally preferred. Carry out the hypothesis test. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion at a 5% significance level.

27. For a newly released movie, executives wanted to investigate a possible relationship between a viewer's age and their rating of a movie. Table 13.15 illustrates the results.

Age / Rating	1⋆	2⋆	3⋆	4⋆
15–30	23	66	112	317
31–45	37	42	87	23
≥ 46	87	102	37	21

Table 13.15: Ratings vs age for movie goers.

The executives wanted to test the null hypothesis that age and rating are independent, and they carried out an appropriate χ^2 test. The value of the appropriate χ^2 test statistic was found to be 377.42. (You should know *how* to calculate this value, but it would be very time consuming to actually calculate it.) Find the appropriate p-value and give an appropriate conclusion.

28. Researchers are investigating the properties of four types of concrete. The four types are subjected to water penetration and cycles of hot and extremely cold temperatures, and the damage (intact, minor damage, major damage) is recorded. The researchers want to determine if the type of concrete has an association (a relationship) with the type of damage. The results of the test are shown in the following table. There are a total of 165 concrete samples.

	Type A	Type B	Type C	Type D
Intact	21	33	10	8
Minor damage	12	10	12	9
Major damage	2	1	22	25

The researchers carry out a chi-square test, and obtain the following output.

```
Pearson's Chi-squared test
data:  concrete
X-squared = 59.6202, df = 6, p-value = 5.376e-11
```

Of the following options, which one of the following is the best conclusion?

(a) There is very strong evidence that the 4 concrete types have the same distribution of damage.
(b) There is little or no evidence of a difference in distribution of damage among the 4 concrete types.
(c) There is very strong evidence of a difference in the distribution of damage among the 4 concrete types.
(d) We can be very confident that concrete type and damage are independent.
(e) There is not strong evidence of a difference in population mean damage between the 4 concrete types.

29. A company is investigating new methods of eradicating bedbugs. They are testing 3 new methods: one based on heat, one based on chemicals, and a combination heat/chemical method. Under controlled conditions, they subject 200 adult bedbugs to each of the 3 methods, and count the number of bedbugs that die. The results are given in Table 13.16.

	Heat	Chemical	Heat and Chemical
Dead	112	54	152
Alive	88	146	48

Table 13.16: Bedbugs killed by different methods.

Suppose we wish to test the null hypothesis that the 3 methods are all equally effective (the true proportion of bedbugs that would be killed under these conditions is the same for all 3 methods). Carry out this test, giving the value of the test statistic, p-value, and conclusion. (Hint to ease the calculation burden: the value of the χ^2 statistic is 97.47.)

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) In words and symbols, what is the null hypothesis in one-way ANOVA?
   (b) In words and symbols, what is the alternative hypothesis in one-way ANOVA?

14.3 Carrying out the One-Way ANOVA

14.3.1 The Formulas

2. Researchers investigated the effectiveness of four different exercise plans on weight loss. Forty people of similar age and other characteristics were randomly assigned to the different plans (10 to each plan). At the end of two months, weight loss was measured. Consider the partially completed ANOVA table found in Table 14.1.

   (a) Complete the ANOVA table (there is enough information to complete the table).
   (b) What is the value of the pooled variance?
   (c) What is the value of the F statistic?

Source	DF	SS	MS	F	p-value
Groups		800			
Error				—	—
Total		1000	—	—	—

Table 14.1: A partially completed ANOVA table.

(d) Is there significant evidence at the 5% level that not all of these plans have the same population mean weight loss?

3. The results from a three group experiment are found in Table 14.2.

	Group 1	Group 2	Group 3
Sample mean	3.1	10.1	18.9
Sample standard deviation	2.1	4.2	3.7
Sample size	10	10	8

Table 14.2: Results from a three-group experiment.

Grand mean: $\bar{X} = 10.11429$.

(a) What is the estimate of the within group variance (s_p^2)?
(b) What is the value of SST?
(c) Complete the entire ANOVA table.
(d) What is the p-value?
(e) Give an appropriate conclusion at the 5% level of significance.

4. Consider the following sample data.

Group A: 4, 6
Group B: 8, 12, 16
Group C: 4, 8

Suppose we wish to analyze this data with one-way ANOVA. (This might not be the best idea, given the very small sample sizes, but suppose we are committed to it.)

(a) What is the value of SST?
(b) What is the value of the pooled variance?
(c) What is the value of the F statistic?

14.3.2 An Example with Full Calculations

5. Walczyk et al. (2013) investigated possible differences between truth tellers and liars when questioned about a mock crime. Participants in a psychology experiment were randomly assigned to a truth telling group, an unrehearsed lying group, or a rehearsed lying group (where the individuals were allowed to see the questions and think about their responses in advance). The participants were then asked several questions. Some questions were general questions, in which all participants were to answer truthfully, and some questions involved a mock crime, in which participants were to either answer truthfully or lie, according to their assigned group. Various characteristics of their responses were recorded.

   In one aspect of the study, the researchers suspected that liars would tend to have greater eye movement when compared to truth tellers. (They also suspected that unrehearsed liars would have greater eye movement than rehearsed liars.) Table 14.3 illustrates the summary statistics of eye movement (in pixels) during the responses to the general questions.

Group	Truth tellers	Unrehearsed lying	Rehearsed lying
Mean	42	63	64
Standard deviation	36	44	38
Sample size	41	42	42

Table 14.3: Summary statistics of eye movement (in pixels).

   (a) Is carrying out a one-way ANOVA appropriate here? How might one investigate the ANOVA assumptions in greater detail?
   (b) Plots of the eye movement measurements (not shown) show a little right skewness. Is it still reasonable to use ANOVA here?
   (c) The following output from R illustrates the results of the one-way ANOVA:

```
             Df Sum Sq Mean Sq F value  Pr(>F)
Groups        2  12757    6378  4.0866 0.01915
Residuals   122 190420    1561
```

   Give a conclusion to the ANOVA F test at the 5% significance level.
   (d) The ANOVA F test showed significant evidence against the null hypothesis at the 5% significance level. Is the analysis complete? What else needs to be done?

14.4 What Should We Do After a One-Way ANOVA?

14.4.1 Introduction

6. Suppose we are about to draw 25 independent random samples and calculate 25 95% confidence intervals. Suppose that for this particular problem the intervals can be considered independent, and the assumptions are perfectly justified in each case. What is the probability that *all* of the intervals capture the parameter that they are estimating?

7. Suppose we are about to draw 10 independent random samples, and use the data to test 10 null hypotheses, each at the 5% significance level. Suppose that for this particular problem the tests can be considered independent, and the assumptions of the test are perfectly justified in each case. What is the probability that there are no Type I errors made in the 10 tests?

14.4.2 Fisher's LSD Method

8. Consider again the study first discussed in Question 5, in which eye movement was measured for truth tellers, unrehearsed liars, and rehearsed liars.

Group	Truth tellers	Unrehearsed lying	Rehearsed lying
Mean	42	63	64
Standard deviation	36	44	38
Sample size	41	42	42

Table 14.4: Summary statistics of eye movement (in pixels).

```
             Df Sum Sq Mean Sq F value  Pr(>F)
Groups        2  12757    6378  4.0866 0.01915
Residuals   122 190420    1561
```

(a) How many pairwise comparisons are there?

(b) Use the LSD procedure to construct confidence intervals for the pairwise differences. Use a confidence level of 95% for each of the individual intervals.

(c) Which intervals contain 0? Which of the observed differences are statistically significant at the 5% level?

(d) What is the family-wise confidence level?

(e) Give a summary of the results.

14.4.3 The Bonferroni Correction

9. Suppose we run an experiment with 5 treatment groups and 6 observations in each group. We carry out a one-way ANOVA, and find a p-value of 0.007. We then wish to calculate confidence intervals for the pairwise comparisons using the Bonferroni multiple comparison procedure.

   (a) How many pairwise comparisons are there?
   (b) If we want to keep the family-wise confidence level at at least 90%, what should be the confidence level of the individual intervals?
   (c) What is the appropriate t value for the calculation of the margin of error of the intervals?

10. Consider again the study first discussed in Question 5, in which eye movement was measured for truth tellers, unrehearsed liars, and rehearsed liars.

Group	Truth tellers	Unrehearsed lying	Rehearsed lying
Mean	42	63	64
Standard deviation	36	44	38
Sample size	41	42	42

Table 14.5: Summary statistics of eye movement (in pixels).

```
             Df Sum Sq Mean Sq F value  Pr(>F)
Groups        2  12757    6378  4.0866 0.01915
Residuals   122 190420    1561
```

   (a) How many pairwise comparisons are there?
   (b) Suppose we wish to use the Bonferroni procedure to construct confidence intervals for the pairwise differences. If we wish to have a family-wise confidence level of 95%, what is the appropriate confidence level for the individual intervals?
   (c) Use the Bonferroni procedure to construct confidence intervals for the pairwise differences. (Helpful tip: Using software we can find that with 122 degrees of freedom, $t_{0.0083} = 2.4274$.)
   (d) Which intervals contain 0? Which of the observed differences are statistically significant at the 5% level?
   (e) How do the Bonferroni intervals compare to the LSD intervals constructed in Question 8?
   (f) Give a summary of the results.

14.4.4 The Tukey Procedure

11. Consider again the study first discussed in Question 5, in which eye movement was measured for truth tellers, unrehearsed liars, and rehearsed liars.

Group	Truth tellers	Unrehearsed lying	Rehearsed lying
Mean	42	63	64
Standard deviation	36	44	38
Sample size	41	42	42

Table 14.6: Summary statistics of eye movement (in pixels).

```
             Df Sum Sq Mean Sq F value  Pr(>F)
Groups        2  12757    6378  4.0866 0.01915
Residuals   122 190420    1561
```

(a) Suppose we wish to use the Tukey procedure to calculate confidence intervals for the pairwise differences. The sample sizes are not equal. Is it still reasonable to use the Tukey procedure?

(b) The results of the Tukey procedure at a family-wise confidence level of 95% are summarized in the following output.

```
      diff        lwr       upr     p adj
TT-RL  -22 -42.579482 -1.420518 0.0331275
UL-RL   -1 -21.455134 19.455134 0.9926103
UL-TT   21   0.420518 41.579482 0.0443732
```

Verify that these are indeed the correct Tukey intervals by calculating them yourself. (Helpful tip: Using software we can find that for 3 groups and 122 degrees of freedom the appropriate Tukey multiplier is 2.3725.)

(c) Which intervals contain 0? Which pairs of means have a statistically significant difference at a family-wise significance level of 5%?

(d) How do the Tukey intervals compare to the LSD intervals constructed in Question 8 and the Bonferroni intervals constructed in Question 10?

(e) Give a summary of the results.

14.5 Examples

14.6 A Few More Points

12. What are the assumptions of one-way ANOVA? What are the consequences if these assumptions are violated?

14.7 Chapter Exercises

14.7.1 Basic Calculations

13. (a) If we let F represent a random variable that has an F distribution with 5 and 8 degrees of freedom, what is $P(F > 4.82)$?
    (b) If we let F represent a random variable that has an F distribution with 3 and 2 degrees of freedom, what is $P(F < 19.16)$?
    (c) If we let F represent a random variable that has an F distribution with 6 and 4 degrees of freedom, what is the value a such that $P(F > a) = 0.05$?
    (d) If we let F represent a random variable that has an F distribution with 12 and 27 degrees of freedom, what is the value a such that $P(F > a) = 0.01$?

14.7.2 Concepts

14. In a one-way ANOVA the F statistic is found to be 0.00002. Does this value give strong evidence against the null hypothesis?

15. Consider a one-way ANOVA.

    (a) Under what conditions would the test statistic be equal to 0?
    (b) Under what conditions would the test statistic be equal to 1?

16. Consider the boxplots in Figure 14.1, representing 3 separate and independent samples of size 20.

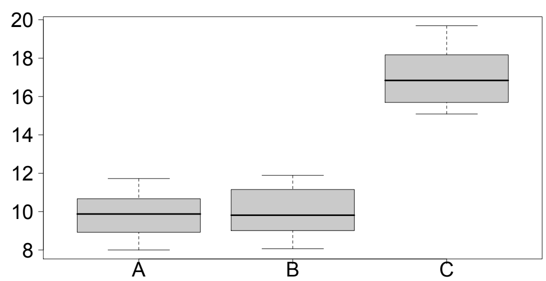


Figure 14.1: Boxplots for 3 independent samples.

Which of the following statements are true?

(a) If we performed a one-way ANOVA on these 3 samples, then the p-value would be small.

(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.
(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.
(d) If we carried out a pooled-variance t-test of H_0: $\mu_A = \mu_C$ against a two-sided alternative, and a one-way ANOVA F test of H_0: $\mu_A = \mu_C$, then the test statistics would have the relationship: $t^2 = F$.
(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.

17. Consider a one-way ANOVA, with 10 observations in each of 5 different groups. Suppose the null hypothesis is true, and the assumptions are true.

    (a) What is the distribution of the test statistic?
    (b) What is the distribution of the p-value?
    (c) What would the p-value equal on average?

18. 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) If the null hypothesis (and the assumptions) are true, then the test statistic in one-way ANOVA has an F distribution.
    (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.
    (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.
    (d) The test statistic in one-way ANOVA can be negative.
    (e) If the null hypothesis is false, then MST will tend to be bigger than MSE.

19. 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) If the null hypothesis is true, then the F statistic will be infinite.
    (b) If the null hypothesis is true, then the F statistic will equal 1.
    (c) If the null hypothesis is false, then the F statistic will sometimes be less than one.
    (d) If the null hypothesis is false, then the F statistic will sometimes be less than 0.
    (e) If the null hypothesis is false, then the p-value will be less than 0.05.

20. 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) If $MSE = 0$, then the p-value will equal the F statistic.
(b) If the population means are equal, then the F statistic will equal 1.
(c) If the sample means are all equal, then the F statistic will equal 1.
(d) If the sample means are all equal, then the null hypothesis is true.
(e) The mean square error is the pooled variance.

21. 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) If the population means are very different, we expect the F statistic to be greater than one.
(b) We will reject the null hypothesis at the 5% level whenever the sample means are all equal.
(c) An assumption of one-way ANOVA is that the groups all have different population variances.
(d) If the assumptions of the pooled-variance t procedures are met, then the t test and the F test are equivalent tests, with $t^2 = F$.
(e) The equal variance assumption is not important for large sample sizes, due to the central limit theorem.

14.7.3 Applications

22. Researchers investigated the effect of calorie-restricted diets on the longevity of mice. They conducted an experiment, randomly assigning each of 331 mice to one of 6 different diets:

- A control group that were allowed to eat an unlimited amount.
- R1, R2, R3, R4, and R5. These were 5 increasingly calorie restricted diets.

The mice were fed the diets until they died. The survival time, in months, was recorded. The results are illustrated in Figure 14.2 and Table 14.7.[1]

Diet	NP	R1	R2	R3	R4	R5
Mean	27.4	32.7	42.3	42.9	39.7	45.1
Standard deviation	6.1	5.1	7.8	6.7	7.0	6.7
Sample size	49	57	71	56	49	49

Table 14.7: Survival time summary statistics for the six diets.

[1] Data from Weindruch, R., Walford, R., Fligiel, S., and Guthrie, D. (1986). The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *The Journal of Nutrition*, 116:641–654. The data used here is simulated data based on the summary statistics from this study.

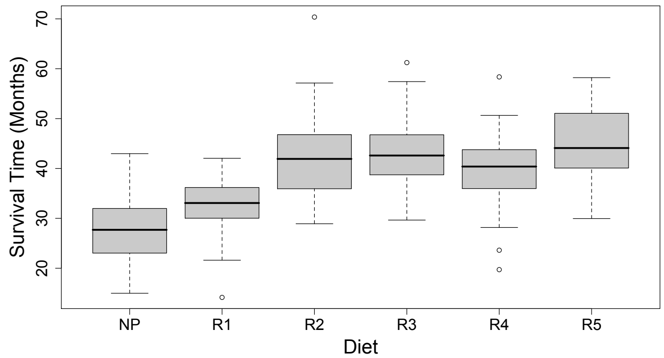


Figure 14.2: Survival times for the different diets.

(a) Do the boxplots and summary statistics give any indication that the ANOVA procedure would be inappropriate? (Do any of the assumptions appear to be violated?)

(b) Consider the following output from the statistical software R:

```
              Df  Sum Sq Mean Sq F value    Pr(>F)
Diets          5 12267.4  2453.5  55.079 < 2.2e-16
Residuals    325 14477.1    44.5
```

Give a proper interpretation of the results of the F test.

(c) There is extremely strong evidence that the groups do not all have the same population mean lifetime. (In other words, there is very strong evidence that the diets do not all have the same effect.) Suppose we wish to investigate the pairwise differences using the LSD procedure. How many pairwise comparisons are there?

(d) Use the LSD procedure to calculate a 95% confidence interval for the difference in means between the NP and R1 diets. Does the interval contain 0? Is the observed difference statistically significant at the 5% level? Give a non-technical summary of the results of this comparison.

(e) Use the LSD procedure to calculate a 95% confidence interval for the difference in means between the R2 and R3 diets. Does the interval contain 0? Is the observed difference statistically significant at the 5% level? Give a non-technical summary of the results of this comparison.

23. Researchers investigated a possible effect of antivenom on swelling after injection of a rattlesnake venom[2]. In the experiment, 24 pigs were injected with *Crotalus atrox* venom in the right hind leg, then they were randomly assigned to one of 3 groups:

- A group that received an intravenous antivenom infusion (IV group).

[2] Offerman et al. (2009). Subcutaneous crotaline Fab antivenom for the treatment of rattlesnake envenomation in a porcine model. *Clinical Toxicology*, 47:61–68.

- A group that received a subcutaneous injection of antivenom at the rattlesnake envenomation site in addition to the IV (SC/IV group).
- A control group that received a saline solution (control group).

Figure 14.3 and Table 14.8 illustrate the change in leg volume after 8 hours.

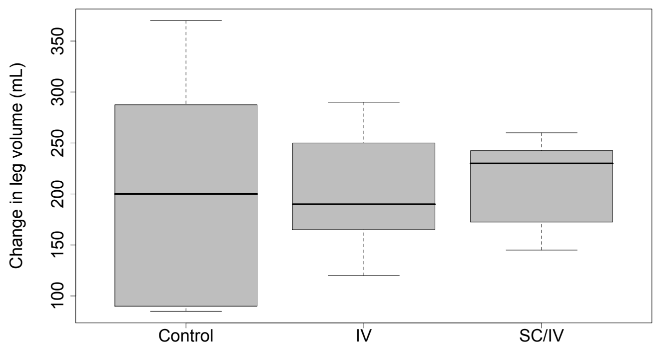


Figure 14.3: Change in leg volume (mL) after 8 hours.

Treatment	Control	IV	SC/IV
Mean	201.25	203.33	209.29
Standard deviation	112.63	56.18	45.04
Sample size	8	9	7

Table 14.8: Summary statistics of change in leg volume (mL) after 8 hours.

(a) Do the boxplots and summary statistics give any indication that the ANOVA procedure would be inappropriate? (Do any of the assumptions appear to be violated?)

(b) Consider the following output for the one-way ANOVA procedure:

```
            Df Sum Sq Mean Sq F value Pr(>F)
Treatments   2    257     128  0.0214 0.9789
Residuals   21 126209    6010
```

The researchers suspected that the antivenom would reduce swelling. Does this experiment give any evidence of this effect?

(c) Are there any other statistical procedures that should be carried out after the ANOVA?

24. A study investigated possible differences in the effect of different therapies on weight gain in young women with anorexia. The young women received either no treatment (Control), a Cognitive Behavioural treatment (CBT), or Family Therapy (FT). The change in weight (kilograms) over the study period was recorded. Figure 14.4 and

Table 14.9 summarize the results.[3]

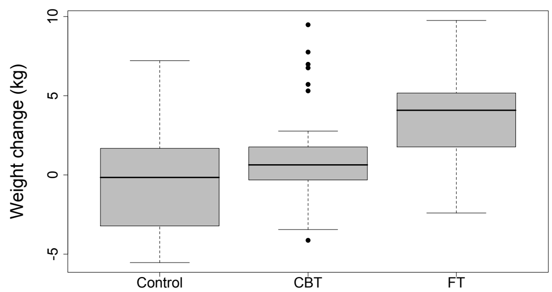


Figure 14.4: Change in body weight (kg) over the course of the study.

Treatment	Control	CBT	FT
Mean	−0.20	1.36	3.30
Standard deviation	3.62	3.32	3.25
Sample size	26	29	17

Table 14.9: Summary statistics of change in body weight (kg) over the course of the study.

(a) Do the boxplots and summary statistics give any indication that the ANOVA procedure would be inappropriate? (Do any of the assumptions appear to be violated?)

(b) Consider the following output from the statistical software R:

```
               Df Sum Sq Mean Sq F value   Pr(>F)
Therapies       2 126.46   63.23  5.4223 0.006499
Residuals      69 804.62   11.66
```

Give an interpretation of the results of the F test.

(c) There is very strong evidence that the groups do not all have the same population mean change in weight. (In other words, there is strong evidence that the therapies do not all have the same effect.) Suppose we wish to investigate the pairwise differences using the LSD procedure. How many pairwise comparisons are there?

(d) Use the LSD procedure to calculate 95% confidence intervals for the pairwise differences in means. Which of the observed differences are statistically significant at the 5% level? Give a non-technical summary of the results of these comparisons.

(e) Use the Bonferroni procedure to calculate confidence intervals for the pairwise differences in means, such that the family-wise confidence level is at least 95%. (Helpful tip: Using software we can find that with 69 degrees of freedom,

[3]The data is taken from: Hand, D., Daly, F., Lunn, A., McConway, K., and Ostrowsky, E. (1994). *A Handbook of Small Data Sets*. Chapman and Hall.

$t_{0.0083} = 2.4537$.) Which of the observed differences are statistically significant? Give a non-technical summary of the results of these comparisons.

25. A study[4] investigated Social Dominance Orientation (SDO) in a male inmate population in the UK. The article states that SDO is a "16-item scale [that] measures a tendency to endorse the appropriateness and desirability of hierarchical relationships." (SDO tends to be high in individuals that believe it is right for certain groups to dominate others.) Is there a difference in the mean SDO of inmates of different ethnic backgrounds?

Ethnicity	White	Asian/Asian British	Black/Black British
Mean	50.89	49.13	53.87
Standard deviation	16.21	14.07	16.80
Sample size	328	23	18

Table 14.10: Summary statistics of SDO by ethnic background.

   (a) Do the values given in Table 14.10 give any indication that ANOVA should not be used? How could the assumptions of ANOVA be further investigated?
   (b) The following output from R shows the results of the ANOVA calculations:

```
               Df Sum Sq Mean Sq F value Pr(>F)
Ethnic groups   2    231     115  0.4439 0.6419
Residuals     366  95077     260
```

   Give a conclusion to the ANOVA F test.

26. Consider again the study first discussed in Question 25, which investigated Social Dominance Orientation (SDO) in a male inmate population in the UK. In Question 25 it was found that there was no evidence of a difference in the mean SDO scores for the different ethnic groups. The author of the paper decided to pool the individuals together (ignoring the ethnic backgrounds) for the subsequent analysis.

   (a) Suppose for a moment that there are meaningful differences in the SDO between the ethnic backgrounds. If these differences are ignored and the inmates are pooled, how might that affect the analysis?
   (b) A major point of interest in the study was investigating possible differences in the SDO scores between different age groups. (Before the study, the author hypothesized that younger inmates would tend to have higher SDO scores.) A summary of SDO scores for different age groups is given in Table 14.11.
   The following output from R shows the results of the ANOVA calculations:

```
            Df Sum Sq Mean Sq F value Pr(>F)
Age groups   3   9265    3088  13.054 3.841e-08
Residuals  391  92497     237
```

[4] Graham-Kevan, N. (2011). Investigating social dominance in a prison population. *Journal of Criminal Psychology*, 1(1):15–23.

Age	18-25	26-35	36-45	45^+
Mean	55.86	51.79	45.23	39.69
Standard deviation	14.90	15.89	15.65	14.67
Sample size	160	133	80	22

Table 14.11: Summary statistics of SDO by age group.

Give a conclusion to the ANOVA F test.

(c) Taking into account the results of the F test, and the summary statistics of Table 14.11, give a non-technical summary of what this study tells us about the relationship between SDO and age group.

14.7.4 Extra Practice Questions

27. Suppose an experiment investigated six different teaching methods. Twenty students were assigned to each of the six teaching methods. After a certain period of time in each program, the students took a test measuring their ability in the subject. The researchers wanted to use one-way ANOVA to analyze the data. The total sum of squares was found to be 8117.2. The best estimate of the within-group variance was found to be 60.2. You have enough information to complete the full ANOVA table.

    (a) Complete the ANOVA table.
    (b) What is the p-value of the test?
    (c) Give an appropriate conclusion at the 5% level of significance.
    (d) If we wished to use the LSD method to calculate 90% confidence intervals for the pairwise differences in the population means, what would be the appropriate margin of error of the intervals?
    (e) What assumptions are necessary for the methods used above to be reasonable? How could we check to see if these assumptions are reasonable?

28. Researchers investigated the strength of 9 different types of seat belt strap. Random samples of four straps of each type were obtained, and the breaking strength, in thousands of Newtons, was measured for each strap. The research points of interest was to test the null hypothesis that the belts all had the same average breaking strength. If this hypothesis was rejected, the researchers wanted to investigate which pairs of means were different. The researchers felt the assumptions of one-way ANOVA are reasonable for this data. Consider the following partially-completed ANOVA table.

Source	DF	SS	MS	F

Treatments Error			12	
Total		612		

(a) Complete the ANOVA table.
(b) What is the p-value of the F test?
(c) Give an appropriate conclusion at the $\alpha = 0.05$ significance level.
(d) Suppose we wish to calculate 95% confidence intervals for the differences in group means ($\mu_1 - \mu_2$, $\mu_1 - \mu_3$, etc.), using the Fisher LSD procedure. What is the appropriate margin of error of the intervals?
(e) Which of the following statements are true?
  i. The sample means were not all equal.
  ii. This was a balanced experimental design (same sample size in each group).
  iii. We know for certain the population mean breaking strengths of the different groups are not all equal.
  iv. The pooled sample variance is equal to 12.

29. An experiment was performed to compare 5 different fuel additives for their effectiveness in improving gas mileage. A total of 20 new cars of the same make and model were used, 4 for each fuel additive. The response variable was the number of kilometres per litre of gas. A completely randomized design was used, and the partially filled out ANOVA table is given in Table 14.12.

Source	DF	SS	MS	F
Treatments Error		680.2		
Total		958.2		

Table 14.12: ANOVA table for the fuel additives question.

(a) In words and symbols, what are the hypotheses of the F test in the ANOVA table?
(b) What assumptions are necessary in order for the one-way ANOVA methods to be reasonable? How could we check to see if these assumptions are reasonable?
(c) Complete the ANOVA table.
(d) What is the p-value of the F test?
(e) Give an appropriate conclusion.

30. Researchers investigated the shear strength of 4 types of tin-lead solder. Each type of solder was used on 5 solder joints (20 joints in total). To investigate possible differences in the mean shear strength between the different types of solder, the researchers carried out a one-way ANOVA in the statistical package R, with the following results.

```
            Df  Sum Sq Mean Sq F value    Pr(>F)
types        3 168.016  56.005  12.657 0.0001709 ***
Residuals   16  70.796   4.425
```

(a) In words and symbols, what are the appropriate hypotheses of the F test in the ANOVA table?

(b) Give an appropriate conclusion to the ANOVA F test.

(c) If we wanted to use Fisher's LSD method to investigate possible differences in the means of the different groups, what is the appropriate 95% margin of error of the individual intervals?

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. What is regression analysis?

2. Suppose we wish to use a student's GPA in high school to help us predict their GPA in first-year university using a regression model.

   (a) What would be the appropriate response variable? What would be the appropriate explanatory variable?

   (b) If we wanted to improve the predictions of university GPA, what other explanatory variables might we consider including in our model?

3. How does *simple* linear regression differ from *multiple* linear regression?

15.2 The Linear Regression Model

4. Consider the simple linear regression model: $Y = \beta_0 + \beta_1 X + \epsilon$. Give interpretations of all of the terms in the model.

5. What is the difference in meaning between $\hat{\beta}_1$ and β_1? What is the difference in meaning between $\hat{\beta}_0$ and β_0?

6. A sample of 30 (X, Y) pairs is drawn, the values are plotted, and a well-fitting line is drawn through the points. The results are illustrated in Figure 15.1. From the plot, estimate the values of $\hat{\beta}_0$ and $\hat{\beta}_1$. (In the future we will use software to calculate these values for us, but we should be able to roughly estimate them from a plot.)

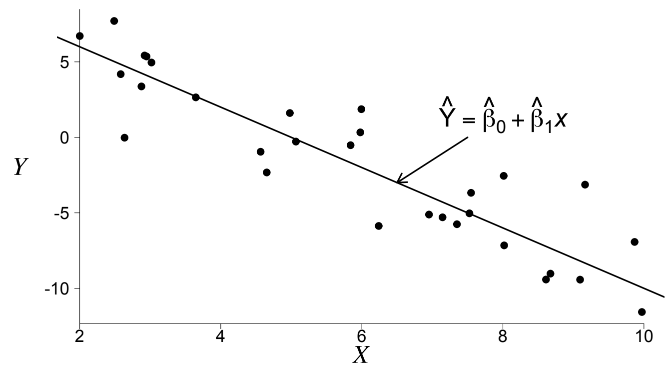


Figure 15.1: A scatterplot with the fitted line $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$. What are the values of $\hat{\beta}_0$ and $\hat{\beta}_1$?

15.3 The Least Squares Regression Model

7. How are the parameter estimates chosen in simple linear regression? What do we call this method of parameter estimation?

8. Suppose in a simple linear regression scenario: $\sum(X_i - \bar{X})^2 = 20.2$, $\sum(Y_i - \bar{Y})^2 = 6.5$, $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 10.1$, $\bar{Y} = 23.0$, and $\bar{X} = 12.0$.

   (a) What are the values of $\hat{\beta}_0$ and $\hat{\beta}_1$?
   (b) If the units of Y are metres, and the units of X are seconds, what are the units of $\hat{\beta}_0$ and $\hat{\beta}_1$?

9. Qu et al. (2011) investigated various relationships between physical characteristics of the lizard *Phrynocephalus frontalis*. In one part of the study, the researchers investigated the relationship between the snout-vent length (mm) and abdomen length (mm) of male lizards. (Part of the study involved investigating how these relationships differed between male and female lizards, but we will focus on just the male lizards here.) The researchers regressed abdomen length (Y) on snout-vent length (X). The results are illustrated in Figure 15.2 and the following output.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.73529    2.50608  -2.289   0.0331
snout_vent   0.63554    0.04847  13.113 2.79e-11
```

Figure 15.2: Abdomen length vs snout-vent length for 22 male lizards (*P. frontalis*).

(a) Based on the output, what is the least squares regression line?
(b) Interpret $\hat{\beta}_1$ in the context of the problem at hand.
(c) Interpret $\hat{\beta}_0$ in the context of the problem at hand.
(d) Based on the regression line, predict the abdomen length for a male *P. frontalis* lizard that has a snout-vent length of 52.0 mm.
(e) One of the lizards in the sample had a snout-vent length of 52.0 mm and an abdomen length of 27.3 mm. What is the residual associated with this lizard?
(f) If we calculated the residuals for all 22 lizards in the sample and added them, what value would we get?

15.4 Statistical Inference in Simple Linear Regression

15.4.1 Model Assumptions

10. What assumptions are required in simple linear regression in order to carry out valid statistical inference procedures?

11. In a simple linear regression analysis with $n = 7$, the residuals were:

$$3.1, 2.4, -6.1, -.5, 8.9, -12.6, 4.8$$

(a) What is the best estimate of σ^2 (the variance about the regression line)?
(b) What is the sum of the residuals? Do they always sum to this value?

15.4.2 Statistical Inference for the Parameter β_1

12. In practice, how often do we carry out a test of:

    (a) H_0: $\hat{\beta}_1 = 0$?
    (b) H_0: $\beta_0 = 0$?
    (c) H_0: $\beta_1 = 0$?
    (d) H_0: $\beta_1 = \beta_0$?

13. What is the sampling distribution of the sample slope? (Under the assumptions of the linear regression model.)

14. Consider again the study first discussed in Question 9, in which researchers investigated the relationship between abdomen length (mm) and snout-vent length (mm) in a type of lizard (*P. frontalis*). The following output summarizes the results of a linear regression of abdomen length against snout-vent length for 22 male lizards.

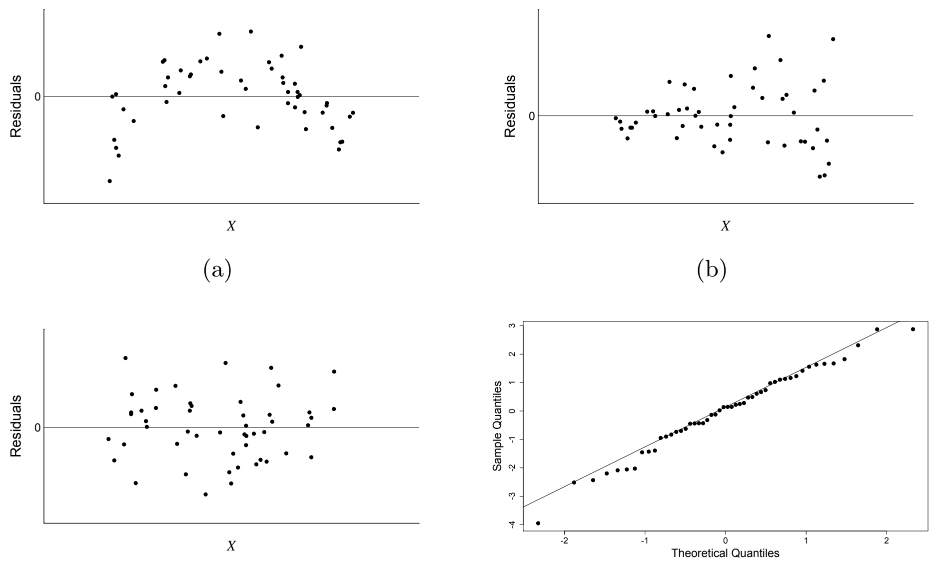
```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.73529    2.50608  -2.289   0.0331
snout_vent   0.63554    0.04847  13.113 2.79e-11
```

    (a) Calculate a 95% confidence interval for β_1.
    (b) Interpret the 95% confidence interval for β_1 found in part 14a.
    (c) Test the null hypothesis that there is no linear relationship between abdomen length and snout-vent length for this type of lizard. Give appropriate hypotheses, test statistic, p-value and conclusion. (Although we would certainly expect the abdomen length to increase with snout-vent length, carry out the test using a two-sided alternative hypothesis.)
    (d) Inference procedures on the intercept β_0 proceed in a similar fashion to inference procedures on β_1, but they are not typically as interesting from a practical viewpoint. Suppose in this question we wish to test H_0: $\beta_0 = 0$ against a two-sided alternative hypothesis. Carry out the test at a 5% significance level.
    (e) Construct a 95% confidence interval for the parameter β_0.

15.5 Checking Model Assumptions with Residual Plots

15. Consider the residual plots given in Figure 15.3. (They represent residual plots for 4 different samples.) Would any of these plots indicate a problem with the assumed linear regression model?

16. Consider the study that related abdominal length (mm) to snout-vent length (mm) for *P. frontalis* lizards (first discussed in Question 9). In Question 14 we carried

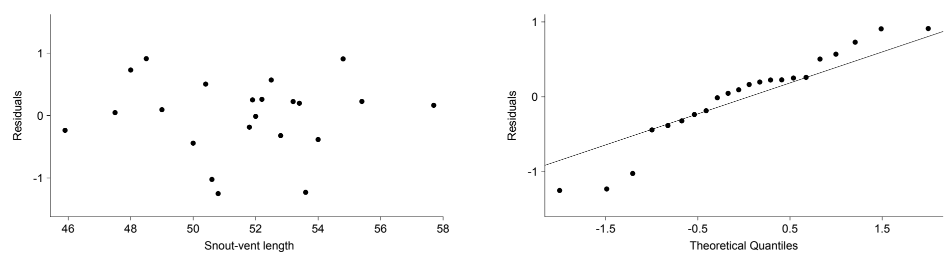


(a) (b)

(c) (d) Normal QQ plot of the residuals.

Figure 15.3: The residual plots for Question 15.

out inference procedures on the slope and intercept. These inference procedures required certain assumptions, and these assumptions should be checked by plotting and looking at appropriate plots of the residuals. Figure 15.4 illustrates a plot of the residuals versus snout-vent length, and a normal quantile-quantile plot of the residuals. Do these residual plots indicate problems with the assumed linear model? Were the inference procedures we conducted in Question 14 valid?



(a) Residuals vs snout-vent length. (b) Normal quantile-quantile plot of the residuals.

Figure 15.4: Residuals vs snout-vent length and a normal quantile-quantile plot of the residuals for Question 16.

15.6 The Correlation Coefficient and the Coefficient of Determination

17. Suppose in a simple linear regression scenario: $\sum(X_i - \bar{X})^2 = 20.2$, $\sum(Y_i - \bar{Y})^2 = 6.5$, $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 10.1$, $\bar{Y} = 23.0$, and $\bar{X} = 12.0$.

    (a) What is the value of r?

    (b) If the units of Y are metres, and the units of X are seconds, what are the units of the correlation coefficient?

18. Consider the plots in Figure 15.5.

    (a) Would the correlation coefficient be a reasonable measure of the strength of the relationship between Y and X in these scenarios? Justify your response.

    (b) Rank the plots in order of increasing values of the correlation coefficient r. (Also take an educated guess as to the value of r in each case.)

    (c) Rank the plots in order of increasing strength of the linear relationship between Y and X.

    (d) The values of Y and X are not given on the axes in Figure 15.5. Would the value of the correlation coefficient change if the scaling were changed on the X and/or Y axes?

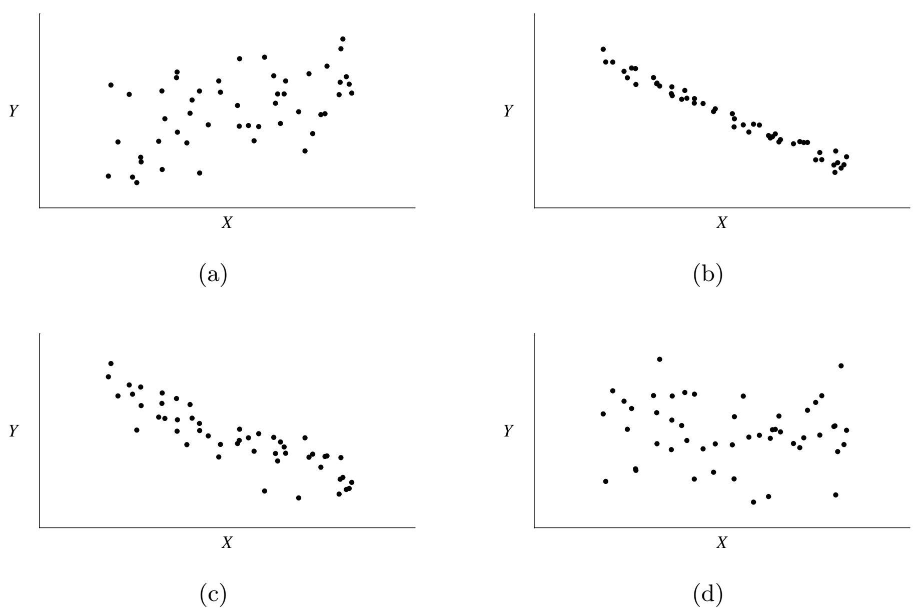


Figure 15.5: The plots for Question 18. Which plot has the largest value of the correlation coefficient r?

19. Consider again the plots in Figure 15.5.

    (a) Would the coefficient of determination R^2 be a reasonable measure of the strength of the relationship between Y and X in these scenarios? Justify your response.
    (b) Rank the plots in order of increasing values of the coefficient of determination R^2. (Also take an educated guess as to the value of R^2 in each case.)

20. Consider again the study first discussed in Question 9, in which researchers investigated the relationship between abdomen length (mm) and snout-vent length (mm) in a type of lizard (*P. frontalis*). The following output summarizes the results of a linear regression of abdomen length against snout-vent length for 22 male lizards. (See Figure 15.2 for the scatterplot.)

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.73529    2.50608  -2.289   0.0331
snout_vent   0.63554    0.04847  13.113 2.79e-11
Residual standard error: 0.6218 on 20 degrees of freedom
Multiple R-squared: 0.8958,       Adjusted R-squared: 0.8906
```

    (a) What is the value of the coefficient of determination in given the output? (Hint: It is called `Multiple R-squared` in the R output.)
    (b) Give an interpretation of the value of R^2 in the context of the problem at hand.
    (c) The correlation coefficient r is not given directly in the output, but it can be found with a simple calculation. What is the value of the correlation coefficient r?

15.7 Estimation and Prediction Using the Fitted Line

21. Consider again the study relating abdomen length (Y) to snout-vent length (X) in a type of lizard (*P. frontalis*). In Question 9 we found that the least squares fit was: $\hat{Y} = -5.73529 + 0.63554X$. The mean snout-vent length for the 22 lizards in the sample is 51.6 mm, and the mean abdomen length is 27.1 mm.

    (a) What is the estimated mean abdomen length for male *P. frontalis* lizards that have a snout-vent length of 50.0 mm?
    (b) What is the predicted abdomen length for a male *P. frontalis* lizard that has a snout-vent length of 50.0 mm?
    (c) Suppose we wish to calculate a 95% confidence interval for the true mean abdomen length for lizards with a snout-vent length of 50.0 mm. Using software,

we can find that the standard error of the estimated mean in this situation is 0.15447. Calculate the appropriate 95% confidence interval.

(d) Suppose we also wish to calculate a 95% confidence interval for the true mean abdomen length for lizards with a snout-vent length of 45.0 mm. Would the appropriate standard error be smaller or larger than the one in part 21c? Would the confidence interval be wider or narrower than the one in part 21c?

(e) Suppose we wished to calculate a 95% prediction interval for the abdomen length of a single lizard with a snout-vent length of 50 mm. Would this prediction interval for a single value be wider, narrower, or the same width as the confidence interval found in part 21c?

15.8 Transformations

15.9 A Complete Example

15.10 Outliers and Influential Points

22. Consider the scatterplot given in Figure 15.6. Four of the points in this data set are plotted in blue and labelled with letters. The regression line on the scatterplot is from a regression with all points included.

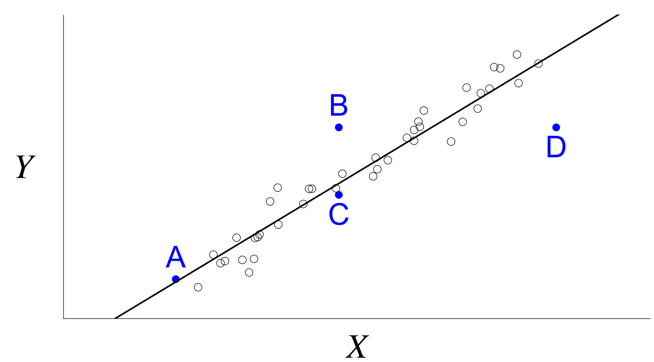


Figure 15.6: A sample data set.

(a) Which two points have the greatest leverage?
(b) Which two points have residuals with the greatest magnitude?
(c) Which point has the greatest influence?
(d) If Point D were removed from the calculations, would the slope of the regression line increase, decrease, or stay the same?

15.11 Some Cautions about Regression and Correlation

15.11.1 Always Plot Your Data

15.11.2 Avoid Extrapolating

23. Consider again the study relating abdomen length (mm) to snout-vent length (mm) in a type of lizard (*P. frontalis*). The results are illustrated in Figure 15.7 and the following output.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.73529    2.50608  -2.289   0.0331
snout_vent   0.63554    0.04847  13.113 2.79e-11
```

Figure 15.7: Abdomen length vs snout-vent length for 22 male lizards (*P. frontalis*).

(a) Would it be reasonable to use the estimated regression line to predict the abdomen length for a lizard with a snout-vent length of 35 mm?

(b) The value of $\hat{\beta}_0$ is given in the output as -5.73529. $\hat{\beta}_0$ is the estimated mean of Y when $X = 0$. But the abdomen length cannot be negative, so this value does not make practical sense. Is our model flawed?

15.11.3 Correlation Does Not Imply Causation

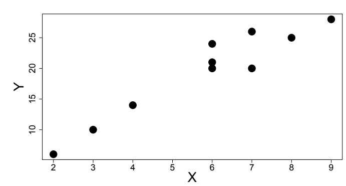
15.12 A Brief Multiple Regression Example

15.13 Chapter Exercises

15.13.1 Basic Calculations

24. Suppose X and Y measurements were taken on 10 individuals, with the results seen in the following table.

Individual	X	Y
1	8	25
2	4	14
3	6	21
4	7	26
5	7	20
6	6	20
7	2	6
8	9	28
9	3	10
10	6	24



Summary statistics: $\bar{X} = 5.8$, $\bar{Y} = 19.4$, $s^2 = 5.15$. The regression line is:

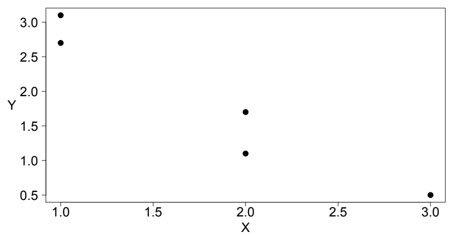
$\hat{Y} = 1.20 + 3.14X$ and $SE(\hat{\beta}_1) = 0.3436$.

(a) What is the value of the residual for the first individual?
(b) What is a 95% confidence interval for β_1?
(c) Based on the 95% interval for β_1, is there strong evidence that X and Y are linearly related?
(d) What is the predicted value of Y if $X = 120$? Is it reasonable to use the regression line for this prediction?
(e) What is the predicted value of Y if $X = 20$? Is it reasonable to use the regression line for this prediction?

25. Consider the following simple data set.

X	1	1	2	2	3
Y	3.1	2.7	1.1	1.7	.5

The following plot and output is a result of a simple linear regression of Y on X for this data.




```
Coefficients:
              Value Std. Error  t value Pr(>|t|)
(Intercept)   4.0571   0.4052   10.0138   0.0021
          x  -1.2429   0.2078   -5.9799   0.0094

Residual standard error: 0.3478 on 3 degrees of freedom
Multiple R-Squared: 0.9226
F-statistic: 35.76 on 1 and 3 degrees of freedom, the p-value is 0.009361
```

(a) If $X = 1$, what is the predicted value of Y based on the regression line?
(b) What is the residual associated with the first value in the data set?
(c) Calculate a 90% confidence interval for the slope parameter.
(d) If there is truly no linear relationship between Y and X, what would the value of the slope parameter be? What would the value of the intercept be?
(e) Suppose we wish to test the null hypothesis that the population slope is equal to 2. What is the value of the appropriate test statistic? What is the p-value of the test? What is an appropriate conclusion?
(f) What assumptions are necessary for the methods used above to be valid? Is there any reason to believe the assumptions are violated in this case?

26. Consider the following R output from a simple linear regression.

```
Coefficients:
             Value  Std. Error  t value Pr(>|t|)
(Intercept) -.6000    1.537       -.390    0.706
          X  1.218     .248       4.919    0.001
```

(a) Is there strong evidence of a linear relationship between X and Y?
(b) If the sample size is 10, what is a 95% confidence interval for β_1?

15.13.2 Concepts

27. What is meant by the term *extrapolation* in simple linear regression?

28. Consider the three plots in Figure 15.8, representing simple linear regressions for 3 different data sets.

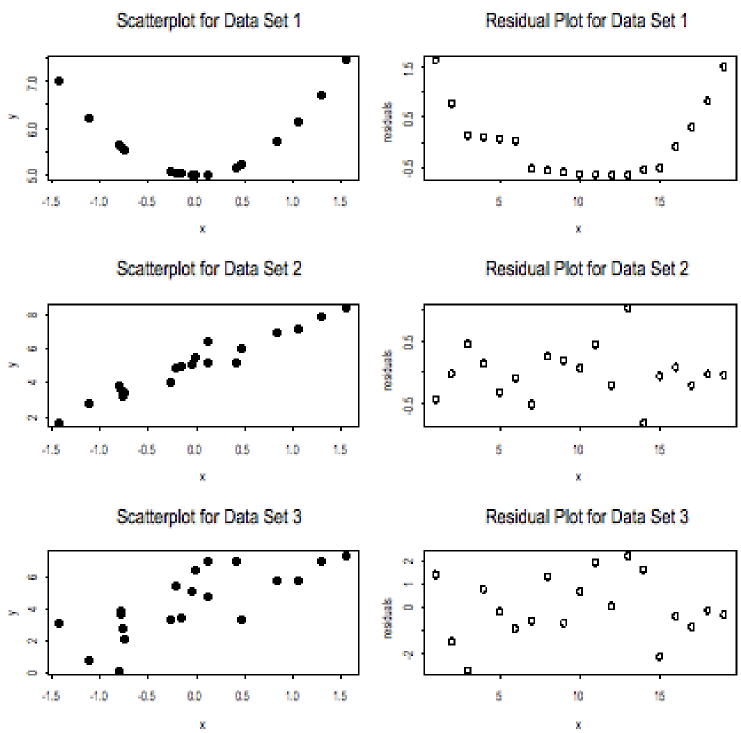


Figure 15.8: Three regression scenarios.

(a) Do any of the residual plots indicate a problem with the assumed model?
(b) For which data set is the strength of the relationship (not necessarily linear) the strongest?
(c) For which data set is the strength of the linear relationship the greatest?
(d) The values of the coefficient of determination are 0.04392, 0.9509 and 0.5935. (Not necessarily in order). Match the data set with its coefficient of determination.

29. If the correlation coefficient is very close to 0, does that imply that there is no relationship between X and Y?

30. If $r^2 = .95$ is there a strong linear relationship between X and Y? Do changes in

X cause changes in Y?

31. The following output from R summarizes the results of a regression analysis with $n = 4$. (It doesn't usually make much sense to try to glean information from a regression with such a small sample size, but this question illustrates the output and some of the calculations.)

```
               Value  Std. Error  t value  Pr(>|t|)
(Intercept)  -1.0000   2.1213     -0.4714   0.6838
          x   5.1000   0.7746      XXXXX    0.0223
```

   (a) What is the estimated regression line?
   (b) Predict the value of Y if $X = 2.5$.
   (c) Calculate a 95% confidence interval for the true slope.
   (d) The test statistic for testing the hypothesis that $\beta_1 = 0$ has been replaced by "XXXXX" out. What is this missing value?
   (e) If we did not have the p-value in the output, and had to look it up in the t table, what would the appropriate range be for the p-value of the test of H_0: $\beta_1 = 0$ against a two-sided alternative?
   (f) The values that were used to calculate the output were:

X	1	2	3	4
Y	4.1	8.2	16.3	18.4

   Calculate the four residuals, and use these residuals to calculate s^2.

32. Suppose in a simple linear regression with $n = 4$, the first 3 residuals are: 2.0, 3.4, -0.7, but you've forgotten to write the fourth residual down. What is the best estimate of σ^2, the variance about the regression line?

33. Suppose we randomly and independently sample 10 X values from the standard normal distribution and 10 Y values from the standard normal distribution. We pair them randomly into 10 (X, Y) pairs, regress Y on X, and test H_0: $\beta_1 = 0$ with a t test.

   (a) What is the distribution of the t test statistic?
   (b) What is the average value of the t test statistic?
   (c) What is the distribution of the p-value in this scenario?
   (d) On average, what would be the value of the p-value?

34. What is the most common hypothesis test in simple linear regression scenarios?

35. Suppose a 95% confidence interval for β_1 is found to be $(1.1, 2.7)$. What does this tell us about the p-value of the test of H_0: $\beta_1 = 0$ against a two-sided alternative hypothesis?

36. 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 method used to find the sample slope and intercept is called the method of least cubes.
    (b) The residuals in a simple linear regression always sum to 0.
    (c) If $\bar{X} = 0$, then $\hat{\beta}_1 = 0$.
    (d) If $\bar{Y} = 0$, then $\hat{\beta}_0 = 0$.
    (e) r has the same sign as $\hat{\beta}_1$

37. 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) $\hat{\beta}_0$ has the same sign as $\hat{\beta}_1$.
    (b) If all data points fall perfectly on a line, then $r = 1$ or -1.
    (c) The least squares regression line always passes through the point (0, 0).
    (d) The least squares regression line always passes through the point $(\bar{X}, \bar{Y})$.
    (e) If there is no linear relationship between Y and X, then $r = 1$.

38. 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) Tests and confidence intervals on $\hat{\beta}_0$ are the most common inference procedures in regression.
    (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.
    (c) $\epsilon_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)$.
    (d) The residuals (the e_i values) are uncorrelated.
    (e) The sample correlation coefficient is the proportion of the variance in Y that is attributable to the linear relationship with X.

39. 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.

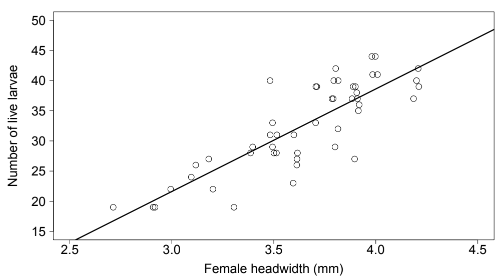
    (a) Instead of least squares, it would be better to choose a regression line based on minimizing the sum of the absolute value of the residuals, but statisticians have not yet figured out how to do so.
    (b) The estimated standard deviation in a simple linear regression is the sum of the squared residuals, divided by the sample size.
    (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.
    (d) If $s = 0$ and $\hat{\beta}_1 > 0$, then $r = 1$.

15.13.3 Applications

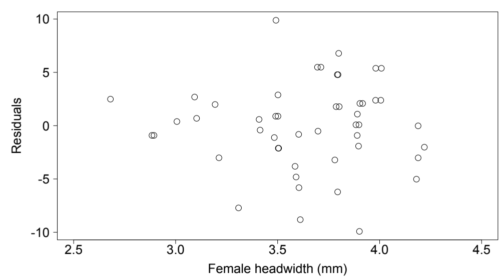
40. Cook et al. (2012)[1] investigated various aspects of larviposition of the blowfly *Calliphora varifrons*. (Larviposition is the laying of larvae (rather than eggs) by insects in which the eggs hatch inside the female.) In one aspect of the study, female fecundity was investigated by trapping female *C. varifrons* blowflies in a field, dissecting them, and counting the number of live larvae they had inside. The researchers wished to investigate whether there was a relationship between the size of the female (as measured by head width, in mm), and the number of live larvae they carried. Figure 15.9 and the following output illustrate the results. (In Figure 15.9 the X values have been *jittererd*—a small random amount has been added to each X. This makes it possible to distinguish two points that lie on top of one another. Jittering is done for visual display only—the non-jittered values are used in the regression calculations.)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -29.400      5.984  -4.913 1.13e-05
headwidth     17.004      1.636  10.396 9.01e-14

Residual standard error: 4.085 on 47 degrees of freedom
Multiple R-squared: 0.6969,        Adjusted R-squared: 0.6905
F-statistic: 108.1 on 1 and 47 DF,  p-value: 9.007e-14
```



(a) Scatterplot number of larvae against head width.



(b) Plot of residuals against head width.

Figure 15.9: Scatterplot of number of larvae against female head width and plot of the residuals.

(a) Do these plots give any indication that the simple linear regression model is not appropriate?

[1]Cook et al. (2012). The laying of live larvae by the blowfly Calliphora varifrons (Diptera: Calliphoridae). *Forensic Science International*, 223:44–66

(b) Can the assumptions of the simple linear regression model be true here? If not, can we still use the model?
(c) What is the estimated relationship between the number of live larvae and female head width?
(d) Give an interpretation of the sample slope $\hat{\beta}_1$ in the context of the problem at hand.
(e) What is the estimate of the difference in the mean number of live larvae carried by female blowflies that differ in their head width by 0.20 mm?
(f) Is there strong evidence of a relationship between the number of live larvae and female head width?
(g) What is the predicted number of live larvae for a female with a head width of 4.0 mm?
(h) What proportion of the variance in the number of live larvae can be explained by the size of the female (as measured by head width)?

41. Fink et al. (2007) investigated a possible relationship between the hand grip strength of young men and their facial attractiveness as perceived by young women. 32 male student volunteers had their hand grip strength measured in kilograms force (kgf), and their facial attractiveness was assessed by 79 female students who viewed pictures of the faces of the men on a computer. The facial attractiveness and hand grip strength for the 32 men are illustrated in Figure 15.10. The following output summarizes the results of the regression calculations.

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.88340    0.94977   1.983   0.0566 .
Grip_strength 0.01309    0.01845   0.709   0.4836

Residual standard error: 0.7952 on 30 degrees of freedom
Multiple R-squared: 0.0165,       Adjusted R-squared: -0.01629
F-statistic: 0.5032 on 1 and 30 DF,  p-value: 0.4836
```

(a) Is there strong evidence of a relationship between hand grip strength and facial attractiveness? Can hand grip strength be used to predict a man's facial attractiveness? Carry out an appropriate hypothesis test, giving hypotheses, the test statistic, p-value, and conclusion.
(b) For this sample data, what percentage of the variance in facial attractiveness can be explained by the linear relationship with hand grip strength?

42. Cámara et al. (2012) investigated various aspects of energy expenditure and lactic acid accumulation in elite cyclists. In one part of the study, the researchers investigated a possible relationship between a cyclist's cadence (measured in rpm) and their Gross Efficiency at the onset of blood lactate accumulation (GE $= \frac{\text{Power output}}{\text{Energy expended}} \times 100\%$). Each of the 12 elite cyclists in the study rode at their

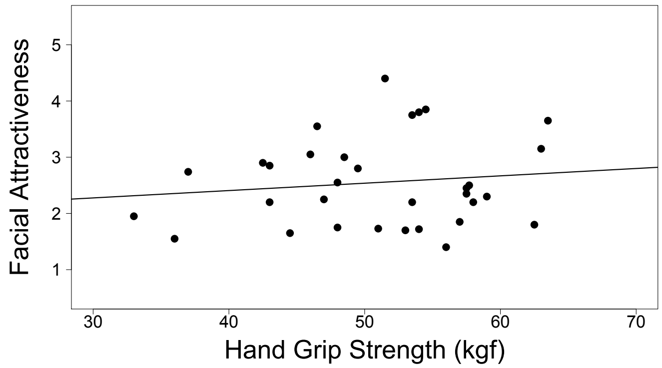
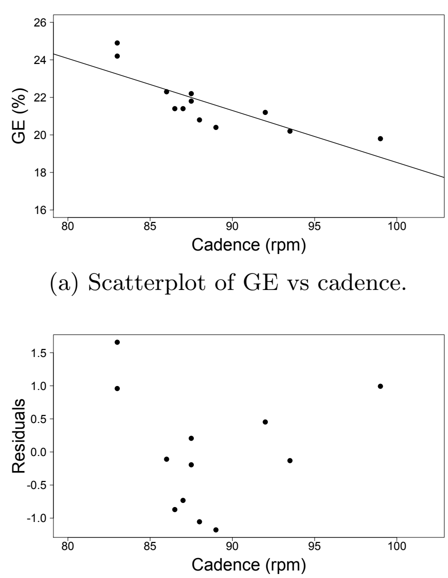


Figure 15.10: Facial attractiveness and hand grip strength for 32 young men.

preferred cadence, and the GE at the onset of blood lactate accumulation was measured. A scatterplot of GE versus cadence with the fitted regression line is given in Figure 15.11a. Figure 15.11b shows the residuals of this fit. The R output for the simple linear regression model is:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 46.24268    5.55656   8.322 8.32e-06
cadence     -0.27713    0.06271  -4.419  0.00130
```



(a) Scatterplot of GE vs cadence.

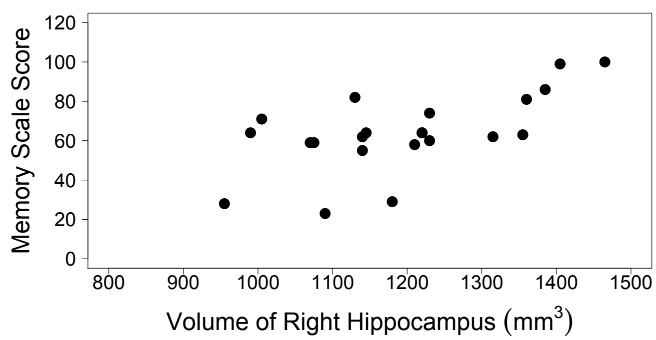
(b) Residuals vs cadence.

Figure 15.11: Scatterplot of GE vs cadence and plot of the residuals for the simple linear regression model.

(a) Do the scatterplot and residual plot show any problems with the assumed linear model? If so, how might we improve the model?

(b) The scatterplot and residual plot give some indication that a line may not adequately model the relationship between cadence and GE. But assume for this question that we feel the simple linear regression model provides a reasonable fit. Do the results of the regression calculations given in the R output indicate strong evidence of a relationship between GE and cadence?

43. Bremner et al. (2007) investigated a possible relationship between memory and the volume of the right hippocampus (an area of the brain) in patients with post-traumatic stress disorder (PTSD). Twenty-two veterans of the Vietnam war had the volume of their right hippocampus measured by an MRI scanner. They were also assigned a score on their verbal memory according to the Wechsler Memory Scale. The following plot and output represent the results of a regression of Memory Scale score on right hippocampus volume.




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -41.72277   29.37690  -1.420  0.17093
hipp_volume   0.08835    0.02440   3.621  0.00170

Residual standard error: 15.7 on 20 degrees of freedom
Multiple R-squared: 0.396,        Adjusted R-squared: 0.3658
F-statistic: 13.11 on 1 and 20 DF,  p-value: 0.001704
```

(a) What are the hypotheses of the hypothesis test that has a p-value of 0.17093? Is this a meaningful test?

(b) What are the hypotheses of the hypothesis test that has a p-value of 0.00170? Is this a meaningful test?

(c) Give a conclusion to the hypothesis test on the slope.

(d) What is a 95% confidence interval for β_1?

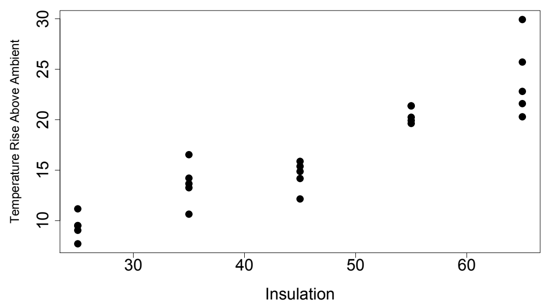
(e) The first observation in the data set had a right hippocampus volume of 955 mm^3 and a Memory Scale score of 28. What is the value of the residual

associated with this value?

(f) What proportion of the variance in memory scores can be explained by the linear relationship with right hippocampus volume?

(g) What is the value of the correlation coefficient?

44. A developer of solar panels is investigating the nature of the relationship between the solar cell temperature rise above ambient (Y) and the amount of insulation (megawatts per square centimetre). They test 25 solar cells under a very specific set of conditions. The following scatterplot and output from the statistical package R illustrate the results.




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.13215    1.51447   0.087    0.931
Insulation   0.36200    0.03211  11.275 7.59e-11
```

(a) What is the predicted value of temperature rise (Y) for an insulation value of 44?

(b) Consider the test in the output has resulted in a p-value of 7.59e-11. Give the hypotheses and an appropriate conclusion to this hypothesis test.

(c) Suppose that before this experiment began, based on theoretical arguments the researchers felt that the slope of the true regression line would be 0.30. Does this experiment provide any evidence against this hypothesis? Test the null hypothesis that the population slope equals 0.30, against a two-sided alternative hypothesis. Give appropriate hypotheses, value of the test statistic, p-value, and conclusion at $\alpha = 0.05$.

45. Suppose a Christmas tree grower wants to know the relationship between the age of a tree and its height, for the species of trees that he grows. A random sample of 10 trees was drawn, and their age and height was recorded.

Some summary statistics:

$n = 10$, $SS_{XX} = 917$, $SS_{YY} = 1519$, $SS_{XY} = 1100$, $s^2 = 24.935$, $\bar{X} = 82$, $\bar{Y} = 260$.

Tree	Age (months) X	Height (cm) Y
1	70	249
2	75	253
3	80	253
...	...	...
10	100	272

You should be able to determine that the least squares regression line is:

$\hat{Y} = 161.6358 + 1.1996X$

(a) When Y is regressed on X, what is the estimate of the slope?
(b) What is the standard error of the slope estimator?
(c) What is a 95% confidence interval for the population slope β_1?
(d) Based on the above interval, is there strong evidence that X and Y are linearly related?
(e) Test the null hypothesis that the population slope is 1, against the alternative that it is different from 1. Give appropriate null and alternative hypotheses, a test statistic, a p-value and a conclusion.
(f) What is the predicted height for a tree that is 92 months old?
(g) What is the estimate of the variance about the regression line?
(h) Given our usual assumptions, what is the sampling distribution of the sample slope?
(i) What is the predicted height of a tree if $X = -2$? Does this prediction make sense in this situation?

Bibliography

Anderson, T., Reid, D., and Beaton, G. (1972). Vitamin C and the common cold: a double-blind trial. *Canadian Medical Association Journal*, 107:503–508.

Brabin et al. (2008). The second to fourth digit ratio (2d:4d) in women with and without human papillomavirus and cervical dysplasia. *American Journal of Human Biology*, 20:337–341.

Bremner et al. (2007). MRI based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *American Journal of Psychiatry*, 152(7):973–981.

Cámara et al. (2012). Influence of pedaling technique on metabolic efficiency in elite cyclists. *Biology of Sport*, 29(3):229–233.

Chakhssi et al. (2010). Change during forensic treatment in psychopathic versus nonpsychopathic offenders. 21:660–682.

Chase et al. (2007). Safety restraint injuries in fatal motor vehicle collisions. *Forensic Science, Medicine, and Pathology*, 3:258–263.

Cook et al. (2012). The laying of live larvae by the blowfly Calliphora varifrons (Diptera: Calliphoridae). *Forensic Science International*, 223:44–66.

Dale, R. (2010). Birth statistics for African (*Loxodonta africana*) and Asian (*Elephas maximus*) elephants in human care: History and implications for elephant welfare. *Zoo Biology*, 29:87–103.

Duran et al. (2013). Growth and weight status of rural texas school youth. *American Journal of Human Biology*, 25:71–77.

Fayrouz et al. (2012). Relation between fingerprints and different blood groups. *Journal of Forensic and Legal Medicine*, 19:18–21.

Fink et al. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology*, 19:82–87.

Franklin et al. (2012). Do ants need to be old and experienced to teach? *The Journal of Experimental Biology*, 215:1287–1292.

Fryar et al. (2012). Anthropometric reference data for children and adults: United states, 2007–2010. *National Center for Health Statistics. Vital Health Stat.*, 11(252).

Graham-Kevan, N. (2011). Investigating social dominance in a prison population. *Journal of Criminal Psychology*, 1(1):15–23.

Guéguen et al. (2012). The sweet smell of. courtship: Effects of pleasant ambient fragrance on womenÕs receptivity to a manÕs courtship request. *Journal of Environmental Psychology*, 32:123–125.

Häkkänen et al. (2009). Gender differences in Finnish homicide offence characteristics. *Forensic Science International*, 186:75–80.

Hand, D., Daly, F., Lunn, A., McConway, K., and Ostrowsky, E. (1994). *A Handbook of Small Data Sets*. Chapman and Hall.

Harlioglu et al. (2012). An investigation on the sperm number and reproductive parameters of males in wild caught freshwater crayfish (*Astacus leptodactylus*, eschscholtz). *Animal Biology*, 62:409–418.

Heidal et al. (2012). Cost and calorie analysis of fast food consumption in college students. *Food and Nutrition Sciences*, 3:942–946.

Holland et al. (2013). A magnetic pulse does not affect homing pigeon navigation: a gps tracking experiment. *The Journal of Experimental Biology*, 216:2192–2200.

Jauchem et al. (2013). Effects of a TASER conducted energy weapon on the circulating red-blood-cell population and other factors in sus scrofa. *Forensic Science, Medicine, and Pathology*, 9:308–320.

Koshy et al. (2010). Parental smoking and increased likelihood of female births. *Annals of Human Biology*, 37(6):789–800.

Kuhn, E., Nie, C., O'Brien, M., Withers, R. L., Wintemute, G., and Hargarten, S. W. (2002). Missing the target: a comparison of buyback and fatality related guns. *Injury Prevention*, 8:143–146.

Leth (2009). Intimate partner homicide. *Forensic Science, Medicine, and Pathology*, 5:199–203.

Lorin de la Grandmaison et al. (2001). Organ weight in 684 adult autopsies: new tables for a Caucasoid population. *Forensic Science International*, 119:149–154.

MacArthur, J. (1931). Linkage studies with the tomato. *Transactions of the Royal Canadian Institute*, 18:1–19.

Mahdavi et al. (2013). Is there any relationship between leb antigen expression and *Helicobacter pylori* infection? *Blood Cells, Molecules and Diseases*, 51:174–176.

Martin et al. (2015). Births: Final data for 2013. National vital statistics reports; vol 64 no 1.

Meinl et al. (2008). Comparison of the validity of three dental methods for the estimation of age at death. *Forensic Science International*, 178:96–105.

Murata, Y. (2001). *Studies on a novel bitter amino acid, pulcherrimine in the green sea urchin gonads*. PhD thesis, Kyoto University.

Offerman et al. (2009). Subcutaneous crotaline Fab antivenom for the treatment of rattlesnake envenomation in a porcine model. *Clinical Toxicology*, 47:61–68.

Petkovšek, S. and Pokorny, B. (2013). Lead and cadmium in mushrooms from the vicinity of two large emission sources in Slovenia. *Science of the Total Environment*, 443:944–954.

Porter et al. (2010). Evaluation of the effectiveness of a brief deception detection training program. *The Journal of Forensic Psychiatry & Psychology*, 21(1):66–76.

Qu et al. (2011). Sexual dimorphism and female reproduction in two sympatric toad-headed lizards, *Phrynocephalus frontalis* and *P. versicolor* (Agamidae). *Animal Biology*, 61:139–151.

Sforza et al. (2009). Age- and sex-related changes in the normal human ear. *Forensic Science International*, 187:110.e1–110.e7.

Shalev (2011). Chronic food restriction augments the reinstatement of extinguished heroin-seeking behavior in rats. *Addiction Biology*, 17:691–693.

Sosa et al. (2005). Life history of *Megamelus scutellaris* with description of immature stages (Hemiptera: Delphacidae). *Annals of the Entomological Society of America*, 98(1):66–72.

Statistics Canada (2013). Canadian health measures survey: Cycle 2 data tables, 2009 to 2011. Statistics Canada Catalogue no. 82-626-X.

Stevenson et al. (2007). Attention deficit/hyperactivity disorder (ADHD) symptoms and digit ratios in a college sample. *American Journal of Human Biology*, 19:41–50.

Tantius et al. (2014). Experimental studies on the tensile properties of human umbilical cords. *Forensic Science International*, 236:16–21.

U.S. Department of Agriculture, Agricultural Research Service (2014). USDA national nutrient database for standard reference, release 27. Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/ba/bhnrc/ndl.

Walczyk et al. (2013). Eye movements and other cognitive cues to rehearsed and unrehearsed deception when interrogated about a mock crime. *Applied Psychology in Criminal Justice*, 29(1):1–22.

Weindruch, R., Walford, R., Fligiel, S., and Guthrie, D. (1986). The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *The Journal of Nutrition*, 116:641–654.

Zorba et al. (2013). Evaluation of the accuracy of different molar teeth measurements in assessing sex. *Forensic Science, Medicine, and Pathology*, 9:13–23.