2. From the same study:

average height of father $\approx$ 68 inches, SD $\approx$ 2.7 inches
average height of son $\approx$ 69 inches, SD $\approx$ 2.7 inches, $r \approx 0.50$

(a) What percentage of the sons were over 6 feet tall?
(b) What percentage of the 6-foot fathers had sons over 6 feet tall?

3. From the same study:

average height of men $\approx$ 68 inches, SD $\approx$ 2.7 inches
average forearm length $\approx$ 18 inches, SD $\approx$ 1 inch, $r \approx 0.80$

(a) What percentage of men have forearms which are 18 inches long, to the nearest inch?
(b) Of the men who are 68 inches tall, what percentage have forearms which are 18 inches long, to the nearest inch?

The answers to these exercises are on p. A65.

6. REVIEW EXERCISES

Review exercises may cover material from previous chapters.

1. The r.m.s. error of the regression line for predicting y from x is ______.
   (i) SD of y (iv) $r \times$ SD of x
   (ii) SD of x (v) $\sqrt{1-r^2} \times$ SD of y
   (iii) $r \times$ SD of y (vi) $\sqrt{1-r^2} \times$ SD of x

2. A computer program is developed to predict the GPA of college freshmen from their high-school GPAs. This program is tried out on a class whose college GPAs are known. The r.m.s. error is 3.12. Is anything wrong? Answer yes or no, and explain.

3. Tuddenham and Snyder obtained the following results for 66 California boys at ages 6 and 18 (the scatter diagram is football-shaped):[9]

   average height at 6 $\approx$ 3 feet 10 inches, SD $\approx$ 1.7 inches,
   average height at 18 $\approx$ 5 feet 10 inches, SD $\approx$ 2.5 inches, $r \approx 0.80$

   (a) Find the r.m.s. error for the regression prediction of height at 18 from height at 6.
   (b) Find the r.m.s. error for the regression prediction of height at 6 from height at 18.

4. A statistical analysis was made of the midterm and final scores in a large course, with the following results:

   average midterm score $\approx$ 50, SD $\approx$ 25
   average final score $\approx$ 55, SD $\approx$ 15, $r \approx 0.60$

   The scatter diagram was football-shaped. For each student, the final score was predicted from the midterm score using the regression line.

(a) For about 1/3 of the students, the prediction for the final score was off by more than ______ points. Options: 6, 9, 12, 15, 25.
(b) Predict the final score for a student whose midterm score was 80.
(c) This prediction is likely to be off by ______ points or so. Options: 6, 9, 12, 15, 25.

Explain your answers.

5. Use the data in exercise 4 to answer the following questions.

   (a) About what percentage of students scored over 80 on the final?
   (b) Of the students who scored 80 on the midterm, about what percentage scored over 80 on the final?

   Explain your answers.

6. In a study of high-school students, a positive correlation was found between hours spent per week doing homework, and scores on standardized achievement tests. The investigators concluded that doing homework helps prepare students for these tests. Does the conclusion follow from the data? Answer yes or no, and explain briefly.

7. The freshmen at a large university are required to take a battery of aptitude tests. Students who score high on the mathematics test also tend to score high on the physics test. On both tests, the average score is 60; the SDs are the same too. The scatter diagram is football-shaped. Of the students who scored about 75 on the mathematics test:

   (i) just about half scored over 75 on the physics test.
   (ii) more than half scored over 75 on the physics test.
   (iii) less than half scored over 75 on the physics test.

   Choose one option and explain.

8. The bends are caused by rapid changes in air pressure, resulting in the formation of nitrogen bubbles in the blood. The symptoms are acute pain, and sometimes paralysis leading to death. In World War II, pilots got the bends during certain battle maneuvers. It was feasible to simulate these conditions in a pressure chamber. As a result, pilot trainees were tested under these conditions once, at the beginning of their training. If they got the bends (only mild cases were induced), they were excluded from the training on the grounds that they were more likely to get the bends under battle conditions. This procedure was severely criticized by the statistician Joe Berkson, and he persuaded the Air Force to replicate the test—that is, repeat it several times for each trainee.

   (a) Why might Berkson have suggested this?
   (b) Give another example where replication is helpful.

9. Every year, baseball's major leagues honor their outstanding first-year players with the title "Rookie of the Year." The overall batting average for the Rookies of the Year is around .290, far above the major league batting average of .260. However, Rookies of the Year don't do so well in their second year—their

6. (Continues exercise 5.) The couples in the previous exercise are followed for a year. Suppose everyone's income goes up by 10%. Find the new regression line for predicting wife's income from husband's income.

7. A statistician is doing a study on a group of undergraduates. On average, these students drink 4 beers a month, with an SD of 8. They eat 4 pizzas a month, with an SD of 4. There is some positive association between beer and pizza, and the regression equation is[13]

   predicted number of beers = _____ × number of pizzas + 2.

   However, the statistician lost the data and forgot the slope of the equation. (Perhaps he had too much beer and pizza.) Can you help him remember the slope? Explain.

8. An investigator wants to use a straight line to predict IQ from lead levels in the blood, for a representative group of children aged 5–9.[14] There is a weak positive association in the data. True or false, and explain—
   (a) He can use many different lines.
   (b) He has to use the regression line.
   (c) Only the regression line has an r.m.s. error.
   (d) Any line he uses will have an r.m.s. error.
   (e) Among all lines, the regression line has the smallest r.m.s. error.

9. In a large study (hypothetical) of the relationship between parental income and the IQs of their children, the following results were obtained:

   average income ≈ \$90,000, SD ≈ \$45,000
   average IQ ≈ 100, SD ≈ 15, $r \approx 0.50$

   For each income group (\$0–\$9999, \$10,000–\$19,999, \$20,000–\$29,999, etc.), the average IQ of children with parental income in that group was calculated and then plotted above the midpoint of the group (\$5,000, \$15,000, \$25,000, etc.). It was found that the points on this graph followed a straight line very closely. The slope of this line (in IQ points per dollar) would be about:

   6,000 3,000 1,500 500 1/500 1/1,500 1/3,000 1/6,000
   can't say from the information given

   Explain briefly.

10. One child in the study referred to in exercise 9 had an IQ of 110, but the information about his parents' income was lost. At \$150,000 the height of the line plotted in exercise 9 corresponds to an IQ of 110. Is \$150,000 a good estimate for the parents' income? Or is the estimate likely to be too high? too low? Explain.

11. (Hypothetical.) A congressional report is discussing the relationship between income of parents and educational attainment of their daughters. Data are

(ii) To win $1 if the first is a king and the second is a queen.

Which option is better? Or are they equivalent? Explain briefly.

3. Four cards will be dealt off the top of a well-shuffled deck. There are two options:

   (i) To win $1 if the first card is a club and the second is a diamond and the third is a heart and the fourth is a spade.
   (ii) To win $1 if the four cards are of four different suits.

   Which option is better? Or are they the same? Explain.

4. A poker hand is dealt. Find the chance that the first four cards are aces and the fifth is a king.

5. One ticket will be drawn at random from the box below. Are color and number independent? Explain.

   | [1] [1] [8] [1] [1] [8] |

6. A deck of cards is shuffled and the top two cards are placed face down on a table. True or false, and explain:

   (a) There is 1 chance in 52 for the first card to be the ace of clubs.
   (b) There is 1 chance in 52 for the second card to be the ace of diamonds.
   (c) The chance of getting the ace of clubs and then the ace of diamonds is $1/52 \times 1/52$.

7. A coin is tossed six times. Two possible sequences of results are

   (i) H T T H T H   (ii) H H H H H H

   (The coin must land H or T in the order given; H = heads, T = tails.) Which of the following is correct? Explain.[7]

   (a) Sequence (i) is more likely.
   (b) Sequence (ii) is more likely.
   (c) Both sequences are equally likely.

8. A die is rolled four times. What is the chance that—

   (a) all the rolls show 3 or more spots?
   (b) none of the rolls show 3 or more spots?
   (c) not all the rolls show 3 or more spots?

9. A die is rolled 10 times. Find the chance of—

   (a) getting 10 sixes.
   (b) not getting 10 sixes.
   (c) all the rolls showing 5 spots or less.

10. Which of the two options is better, or are they the same? Explain briefly.

    (i) You toss a coin 100 times. On each toss, if the coin lands heads, you win $1. If it lands tails, you lose $1.
    (ii) You draw 100 times at random with replacement from | [1] [0] |. On each draw, you are paid (in dollars) the number on the ticket.

11. In the box shown below, each ticket should have two numbers:

| [1 |] [1 | 2] [1 | 2] [1 | 3] [3 | 1] [3 | 2] [3 |] [3 |] |

A ticket will be drawn at random. Can you fill in the blanks so the two numbers are independent?

12. You are thinking about playing a lottery. The rules: you buy a ticket, choose 3 different numbers from 1 to 100, and write them on the ticket. The lottery has a box with 100 balls numbered from 1 through 100. Three balls are drawn at random without replacement. If the numbers on these balls are the same as the numbers on your ticket, you win. (Order doesn't matter.) If you decide to play, what is your chance of winning?

7. SUMMARY

1. The *frequency theory* of chance applies most directly to chance processes which can be repeated over and over again, independently and under the same conditions.

2. The chance of something gives the percentage of times the thing is expected to happen, when the basic process is repeated over and over again.

3. Chances are between 0% and 100%. Impossibility is represented by 0%, certainty by 100%.

4. The chance of something equals 100% minus the chance of the opposite thing.

5. The chance that two things will both happen equals the chance that the first will happen, multiplied by the *conditional* chance that the second will happen given that the first has happened. This is the *multiplication rule*.

6. Two things are *independent* if the chances for the second one stay the same no matter how the first one turns out.

7. If two things are independent, the chance that both will happen equals the product of their unconditional chances. This is a special case of the multiplication rule.

8. When you draw at random, all the tickets in the box have the same chance to be picked. Draws made at random with replacement are independent. Without replacement, the draws are dependent.

9. Blindly multiplying chances can make real trouble. Check for independence, or use conditional chances.

10. The mathematical theory of chance only applies in some situations. Using it elsewhere can lead to ridiculous results.