Answers to Review Exercises

Part I. Design of Experiments

Chapter 2. Observational Studies

1. (a) Too hasty. What about population size?

   Comments: Michigan may include the big bad city, but Minnesota has twice the population of Michigan. The crime rate is lower in Michigan. (Of course, there probably are neighborhoods in Detroit that are best avoided.)

   (b) This is better reasoning, because the population of the U.S. increased over the period 1991–2001. Looking at rates would make the point even more clearly. (In fact, there has been a remarkable decline in crime rates over the period 1980–2005.)

2. (a) What's missing is the number of cars on the road, and there were a lot more Corvettes. (For example, 33,586 Corvettes were sold in 2002, versus 8,065 Q45s.) You need to look at rates. If anything, thieves prefer the Q45—a much classier car.

   (b) Same issue as (a).

   (c) False. The rate is low because the denominator is large relative to the numerator. The rate compares the number of jeeps stolen to the number sold. That's the point of using a rate.

3. No. In the Salk trial, the parents who consented were on the whole better off than the parents who did not consent, and their children were more at risk to begin with (p. 4).

4. (a) They were controlling for age and sex as possible confounders; this is discussed on p. 13, with respect to a specific disease—lung cancer.

   (b) This is the wrong conclusion to draw. Ex-smokers are a self-selected group, and many people give up smoking because they are sick. So recent ex-smokers include a lot of sick people. (Other epidemiological data show that if you quit smoking, you will live longer.)

5. No. The data from the double-blind study are more reliable, and suggest that the results from the single-blind were biased.

6. Subjects who did not improve during the first part of the trial probably concluded that they were on the placebo (whether they were or they weren't) and would be switched to the "real" medication during the second part of the trial. This expectation made them improve—the placebo effect.

7. (a) This is an observational study, so confounding may be a problem.
   (b) Rates of cervical cancer go up with age. Women of different marital status have different patterns of sexual activity, and are therefore exposed to different kinds of risk; similarly for education. In other words, age, marital status, and education are potential confounders.
   (c) Pill users are more active sexually than non-users, and have more partners. That seems to be what makes the rate of cervical cancer higher among pill users. (This is like example 2 on p. 16 or exercise 11 on p. 23.)
   (d) No; see (c).

8. Memorial Day is at the end of May; Labor Day is early in September. Just over 25% of the days of the year fall in between. Even if burglars work the same amount every day, over 25% of the burglaries would occur between Memorial Day and Labor Day.

9. (a) False. (b) True. (c) False: that is the whole point of experiments.

   Discussion. People who eat lots of fruits and vegetables are different from the rest of us in many other ways. Some other aspect of diet or life style may be protective. Of course, the observational studies might be right; something in the fruits and vegetables other than the vitamins might be the protective factor.

10. (a) Observational study. (b) Yes. (c) Yes.
    (d) No. The gene would also have to be associated with controlling behavior by the mother (p. 20).
    (e) A mother who sees her child eat too much might respond in a way that psychologists would interpret as "controlling"—Johnny, stop eating!
    (f) No. The *Chronicle* seems to have over-reacted.

11. (a) The treatment group consists of those who finished boot camp. The control group consists of other prisoners—including those who do not volunteer, or those who volunteer but do not complete the program.
    (b) This is observational. The prisoners decide whether to volunteer for boot camp and whether to stay in the program or drop out. That is the problem: those who volunteer and stay the course might be quite different from who volunteer but drop out.
    (c) False.

    Comment. An experiment could be done either like the polio trial (p. 1ff) or the HIP trial (exercise 9 on pp. 22–23):

    Like the polio trial. Ask for volunteers. Randomize some of the volunteers to treatment (assignment to boot camp) and some to control. Compare the recidivism rate for the two groups—but include the dropouts in the treatment group. (Otherwise, you still have the problem of self-selection.)

Like the HIP trial. Take a group of prisoners. Randomize some to treatment (invitation to participate in boot camp) and some to control. Compare the recidivism rate for the two groups—but include in the treatment group those who decline to participate and those who drop out. (Again, this is to guard against the problem of self-selection.)

12. False. The conclusion does not follow. This is just like the admissions study (pp. 17ff). The Democrats may be concentrated in wards with low turnouts. Here is an example, with only two wards (and see exercise 13 on p. 24).

	DEMOCRATS		REPUBLICANS	
	Total number	Number voting	Total number	Number voting
Ward A	1000	100	100	5
Ward B	100	60	1000	500

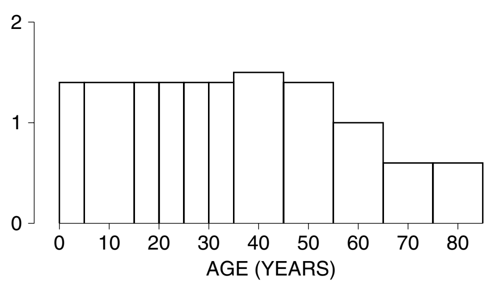
Part II. Descriptive Statistics

Chapter 3. The Histogram

1. 66 inches, 72 inches.

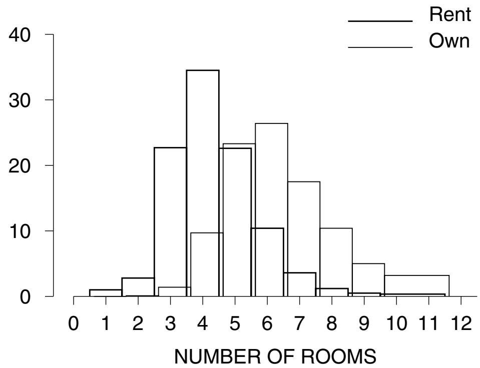
2. (a) There are more at age 1. The histogram is higher at 1 than at 71.
   (b) There are more at age 21.
   (c) There are more age 0–4,
   (d) 50%

Histogram for review exercise 2, chapter 3

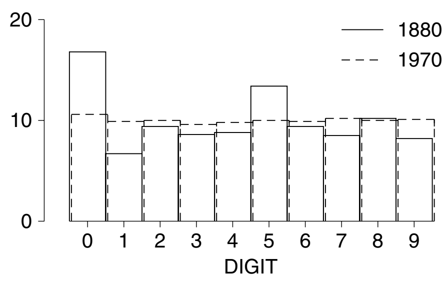


3. (a) Rounding.
   (b) No. Taking percentages adjusts for the difference between the total numbers. On the whole, rental units tend to be smaller.
   (c) Rental units are smaller, as noted above.

Histograms for review exercise 3, chapter 3
(The histogram for owners is shifted to the right)



4. (a) 25% (b) 99% (c) 140–150 mm (d) 135–140 mm
   (e) About $5 \times 2.1 = 10.5\%$ (f) 102–103 mm
   (g) 117–118 mm is a good guess; the interval is somewhere between 115 and 120 mm.
5. \$10 thousand $\times$ 1% per thousand dollars = 10%.
6. (i) and (ii) not (iii).
   Reason: With lists (i) and (ii), 25% of the people have heights between 66.5 inches and 67.5 inches; 50% between 67.5 and 68.5 inches; 25% between 68.5 and 69.5 inches. Not so with list (iii).
7. (i) Natural causes. (ii) Trauma.
   Reason: Young people die of accidents, murder, etc. Old people die of heart disease, cancer, etc.
8. (a) is true, (b) and (c) are false. The figure does not adjust for the different lengths of the class intervals, and is misleading for that reason.
9. (a) True. (b) True.
   (c) People with failing GPAs may round them up; and 2 is such an important number—for GPAs—that people with GPAs just above 2 may round them down.
10. (a) The histogram is shown below.



10. (b) In 1880, people did not know their ages at all accurately, and rounded off.
    (c) In 1970, people knew when they were born.
    (d) In 1880, there was a strong preference for even digits (although 4 and 6 lose out to 5); again, this is probably due to rounding. In 1970, the preference was much weaker.
11. The lowest of the top 15 scores is 90. Then there is a gap of 6 points—the next score is 84. There is no gap anything like this big in the rest of the distribution.
12. False. There are very few days where the temperature is above 90 degrees. The investigators should have looked at the number of riots divided by the number of days in each temperature range.

Chapter 4. The Average and the Standard Deviation

1. (a) Average = 50. Deviations = −9, −2, 0, 0, 4, 7. The SD is

$$\sqrt{\frac{(-9)^2 + (-2)^2 + 0^2 + 0^2 + 4^2 + 7^2}{6}} = 5.$$

   (b) 48, 50, 50 are within 0.5 SDs of average, i.e., in the range from 47.5 to 52.5.

   48, 50, 50, 54, 57 are within 1.5 SDs of average, i.e., in the range from 42.5 to 57.5.
2. (a) List (ii) has the smaller SD, because it has more entries at the average.
   (b) This time, list (i) has the smaller SD: list (ii) has two wild entries, 1 and 99.
3. (a) 5. The average should be in the middle of the distribution: only three of the numbers are smaller than 1, and none are bigger than 10.
   (b) 3. If the SD is 1, the entries 0.6 and 9.9 are much too far from average. The SD can't be 6, because none of the numbers are more than 6 away from the average.
4. For income, the average will be bigger than the median—long right hand tail at work (pp. 64–65). For education, the average will be smaller than the median: the histogram has a long left hand tail (p. 39).
5. 80 mm is more than 3 SDs below average, and is unusually low. 115 mm and 120 mm are about average. 210 mm is unusually high.
6. (a) (i) 60 (ii) 50 (iii) 40
   (b) (i) median is bigger than the average—long left hand tail
   (ii) median is about equal to the average—symmetry
   (iii) median is less than the average—long right hand tail
   (c) 15. Most of the area is within 50 of the average, so 50 is too big. Only a little of the area is within 5 of the average, so 5 is too small.
   (d) False. Histograms (i) and (iii) are almost mirror images, and have just about the same SD.

7. (a) Average weight of men = $66 \times 2.2 \approx 145$ pounds, SD ≈ 20 pounds. Average weight of women ≈ 121 pounds, SD ≈ 20 pounds.

   (b) 68%: the range is average $\pm$ 1 SD.

   (c) bigger than 9 kg: if you take the men and the women together, the spread in weights goes up.

8. (a) The girls have to be the same height (on average) as the boys at age 11. Otherwise, the average height of all the children would differ from the average of the boys.

   (b) $(137 + 151)/2 = 144$ cm.

9. (a) Yes: the average goes up by ($986,000 − $98,600)/1000 = $887.40.

   (b) No. (That is one advantage of the median: it is not thrown off by outliers.)

10. (a) 163, the average.

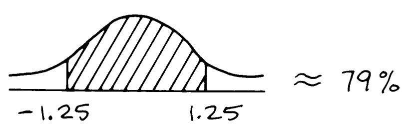
    (b) $8, because the SD is 8; see p. 68.

11. $8, because the SD is 8, and the SD is the r.m.s. deviation from average.

12. The data are cross sectional not longitudinal, so the data only provide weak support for the theory. (Longitudinal data show that most spells of poverty are short; see the references in note 15 to chapter 4.)

Chapter 5. The Normal Approximation for Data

1. (a) 79% of 25 $\approx$ 20.

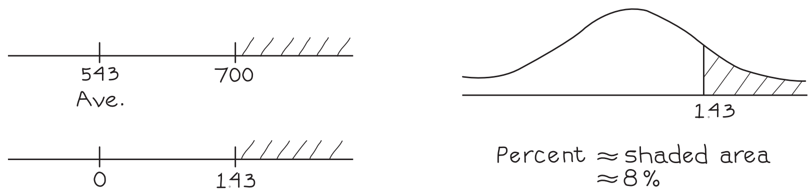


   (b) 18.

2. Something is wrong with the computer. The entries should be around 1 in size, but they are way too big. The first score, for example, is 6.2 SDs below average; the eighth score is even farther down.

   Comment for mathematicians. A list of 100 numbers cannot have an entry more than $\sqrt{99} \approx 10$ SDs away from average. *Sketch of proof*. (i) Standarize the list to have mean 0 and variance 1. (ii) If the largest entry is L, minimize the sum of squares of the other 99 entries, their sum being constrained to equal $-L$.

3. (a) In 1967, a score of 700 was 1.43 SDs above average: $(700 - 543)/110 \approx 1.43$. The percentage scoring over 700 was about 8%—



(b) In 1994, a score of 700 was 1.83 SDs above average, and the percentage scoring over 700 was about 3%.

4. (a) 9% (b) 5%

5. The percentage of men with heights between 66 inches and 72 inches is exactly equal to the area between 66 inches and 72 inches under the histogram. This percentage is approximately equal to the area between -1 and $+1$ under the normal curve.

6. No. For example, the normal curve says that about 16% of the scores should be more than 1 SD above average, and none are.

7. (a) 400 is 1.5 SDs below average; this student is in the 7th percentile of the score distribution (approximately), because the area to left of -1.5 under the normal curve is 7%.

   (b) The student needs to be about 0.7 SDs above average, so, needs a score of about 620 on the M-SAT. (The 75th percentile is nearly 0.675, rounded here to 0.7.)

8. (a) True. See pp. 92ff.

   (b) False: all the deviations from average stay the same.

   (c) True.

   (d) True: all the deviations from average are doubled.

   (e) True.

   (f) False: all the deviations from average have their signs changed, but that goes away in the squaring. The SD has to be positive (or, exceptionally, zero).

9. (a) False. For example, the list 1, 2, 99 has a median of 2 and an average of 34.

   (b) False. For the list 1, 2, 99, two-thirds of the entries are below average.

   (c) False. For example, income data (p. 88) have a long right-hand tail; educational levels have a long left-hand tail as well as bumps at 8, 12 and 16 years (p. 39).

   (d) False. If the histogram for a list follows the normal curve, the percentage of entries within 1 SD of average will be around 68%. Otherwise, the percentage could be a lot different. For instance, take a list of eight numbers: one is 30, one is 70, and six are 50. The average is 50, the SD is 10, and 75% are within 1 SD of average.

10. The histogram has a long right hand tail, the median is well below average. Therefore, a lot less than 50% are earning above \$32,000. Choose 40%.

11. Histogram (i) is right. With (ii), the average of 1.1 would be right in the middle, and a lot of people would be taking fewer than $1.1 - 1.5 = -0.4$ courses. Histogram (iii) is even worse.

12. The students who take the subject-matter tests are the ones applying to the better schools—and in general, these are the better students.

Chapter 6. Measurement Error

1. False: each measurement is thrown off by chance error, and this changes from measurement to measurement. (It is a good idea to replicate measurements, so as to judge the likely size of the chance error.)

2. (a) The tape may have stretched; the hook at the end may have moved.
   (b) Cloth.
   (c) Yes, as the tape stretches or the hook moves.

3. (a) False. Chance errors are sometimes positive and sometimes negative. Bias pushes in one direction.
   (b) False, same reason.
   (c) True.

4. 0.03 inches or so.

5. (a) No. Persons 2 and 10 copied from each other: they got exactly the same answers, with the decimal point in the wrong place. The other students probably worked independently.
   (b) First, nobody got the same answer both times. Second, there is a lot of person-to-person variation.

Chapter 6. Special Review Exercises

1. False; see p. 89.

2. (a) All the numbers on the list are the same. Example: 2, 2, 2.
   (b) All the numbers on the list are 0.

3. We can get the SD by looking at the first two scores, in original units and in standard units: $79 - 64 = 15$ points, while $1.8 - 0.8 = 1.0$. So, 1.0 SD = 15 points, and the SD is 15 points. Next, we get the average from the fact that 64 is 0.8 in standard units: 0.8 SD = 12 points, so the average must be $64 - 12 = 52$. Finally, we complete the table:

52	72	31
0	1.33	−1.4

4. (a) Approximately equal to the area under the curve between −1.5 and 1.5, or 87%.

(b) 560. The range 450–650 corresponds to ± 1 in standard units. About 68% of the students at the university had scores in this range. There must have been about $1000/0.68 \approx 1470$ students at the university. The range 500–600 corresponds to ± 0.5 in standard units. About 38% of the students at the university had scores in this range: 38% of $1470 = 0.38 \times 1470 \approx 560$. Moral: the normal curve is not a rectangle.

5. Probably a recording error was made at one interview or the other.

6. (a) $(600 + 650)/2 = 625$

   (b) More than 125. When you put the men and women together, the spread goes up, because the two distributions are different.

7. The average is 630; again, the SD is more than 125.

8. You would be too low, because the curve is lower than the histogram in that range.

9. Average $= 10 - 6.4 = 3.6$ and SD $= 2.0$.
   Reason: number wrong $= 10 -$ number right.

10. False. The data are cross sectional not longitudinal. The 20-year-olds were born 1956–60; the 70-year-olds, 1906–10. In the early part of the century, there was much more pressure to conform and be right-handed.

    Comment. Some investigators believe that left-handed people suffer higher mortality, and use the trend in left-handedness as supporting evidence. See Stanley Coren, "Left-handedness and accident-related injury risk" *American Journal of Public Health* vol. 79 (1989) pp. 1040–41, and "The diminished number of older left-handers: differential mortality or social-historical trend?" *International Journal of Neuroscience* vol. 75 (1994) pp. 1–8. However, Coren's argument seems a little shaky. For some evidence in the other direction, see M. E. Salive et al., "Left-handedness and mortality," *American Journal of Public Health* vol. 83 (1993) pp. 265–7.

11. By symmetry, the average is half way between the 25th and 75th percentiles, and is 64.0 inches. The 75th percentile is about 0.67 SDs above average, so, the SD is $(65.8 - 64.0)/0.675 \approx 2.7$ inches. The 90th percentile is about 1.28 SDs above average, that is, $64.0 + 1.28 \times 2.7 \approx 67.5$ inches.

12. The likely explanation is digit preference. People round their incomes to the nearest \$1,000 or \$10,000. The class interval \$15,000–\$17,500 includes the left endpoint (a beautiful round number) as well as \$16,000 and \$17,000. The next class interval, \$17,500–\$20,000, includes \$18,000 and \$19,000 but not \$20,000: class intervals include left endpoints not right endpoints. So this second interval has only two round numbers in it, rather than three; neither is as round as \$15,000. The other pairs of intervals are similar to these two. People with incomes below \$15,000 seem less prone to rounding: every dollar counts.

13. (a) Heart disease rates go up with age: the drivers could be older, accounting for the difference in rates. That explanation has been eliminated.

(b) You want the difference in rates to be due to exercise on the job, rather than pre-existing factors. It might take some time for exercise to have its beneficial effect.

(c) Drivers and conductors are going to be similar with respect to age, education, income, and so forth.

(d) This is an observational study, so there may be some confounding. Drivers may be more at risk than conductors to start with. For example, drivers may be heavier.

(e) You could look to see if the drivers and conductors had similar body sizes when they were hired; that would suggest the two groups were comparable at time of hire, and strengthen the argument that exercise matters. If drivers were heavier than conductors at time of hire, the argument for exercise is weaker.

Comment. As it turned out, the drivers were heavier than the conductors at time of hire, explaining the difference in rates of heart disease.

14. Moving high-risk women from the control group to the treatment group lowers the death rate in the control group and increases the death rate in the treatment group. That biases the study against screening.

The assignment does increase the detection rate in the treatment group, but that is not how benefits are measured. The number of lives saved has to be determined by comparing the death rates in the treatment and control groups, and the bias runs against treatment. (Doctors can seldom tell what would have happened if a patient had chosen a different treatment: that is why clinical trials are needed.)

15. The experimental comparison is treatment (groups B + C) versus controls (group A). The investigator is making observational comparisons, because treatment cases select themselves into group B or C. Cases that opt for a pretrial conference are probably different from cases that don't—so there will be lots of confounding. Similar issues came up when comparing the consent and no-consent groups in the Salk vaccine field trial (section 1.1), the adherers and the non-adherers in the clofibrate trial (section 2.2), or the examined and refused groups in the HIP trial (exercise 9 on pp. 22–23). Furthermore, the investigator does not seem to be presenting all the data. For example, there were 2954 cases in all; only 2780 are reported in the first table. Also, 22% of 701 $\approx$ 154 of the group B cases reached trial (first table). The second table only reports on 63 cases in group B. There is a similar issue for groups A and C.

Part III. Correlation and Regression

Chapter 8. Correlation

1. The answer is (d). With (a), the averages are too low. With (b), the SDs are too small. With (c), the SDs are too big and the correlation is too high.

2. (a) Negative: older cars get fewer miles per gallon.

(b) Richer people own newer cars, and maintain them better. (Some rich people own Ferrari gas guzzlers, but not many; and 10-year-old Chevrolets in poor repair might guzzle even more.)

3. The correlation would be 1.00. All the points on the scatter diagram for height of wife vs. height of husband would lie on a straight line which slopes up. The slope of the line is 0.92, but correlation and slope are two different things.

4. 0.3. Taller men do marry taller women, on average. But there is lots of variation around the line.

5. (i) 0.60 (ii) 0.30 (iii) 0.95

   Reasoning: correlation (iii) must be nearly 1; correlations (i) and (ii) are moderate, with (i) being stronger.

6. This is false (p. 126).

7.

0.62	−1.00
−0.85	0.97
0.06	−0.38

8. (a) 42 inches (b) 2.5 inches (c) 0.80 (d) solid

9. (a) −0.80 (b) 0.3 (c) 1.00

   Comment. In (c), all the points lie on the line $y = 2x$, so there is no need to do any arithmetic.

10. Smaller when the score on form L is 75. If you take narrow vertical chimneys over 75 and 125 in the scatter diagram, there is less vertical spread—therefore, less uncertainty in the predictions—in the chimney over 75.

11. $r = -1$: number wrong $= 10 -$ number right, so all the points on a scatter diagram (for number wrong vs. number right) lie on a straight line which slopes down. Also see exercise 9 on p. 106.

12. (a) Three students got 91 on the first count and 82 on the second, so they probably worked together. Another three got 85 on both counts; however, since that is the right answer, these students are probably just good counters.

    (b) False.

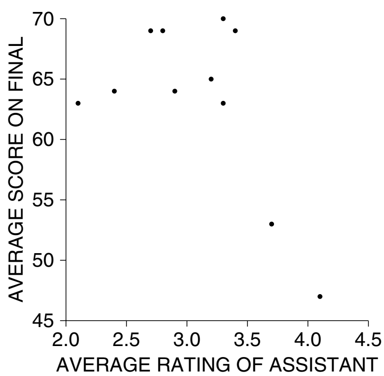
Chapter 9. More About Correlation

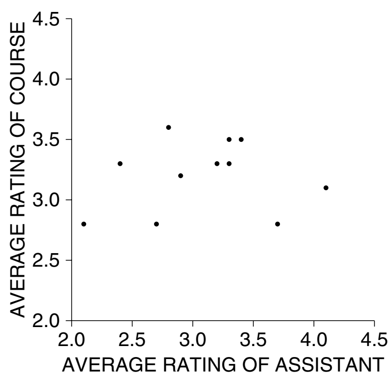
1. When studying one variable, you can use a graph called a histogram. When studying the relationship between two variables, you can use a graph called a scatter diagram.

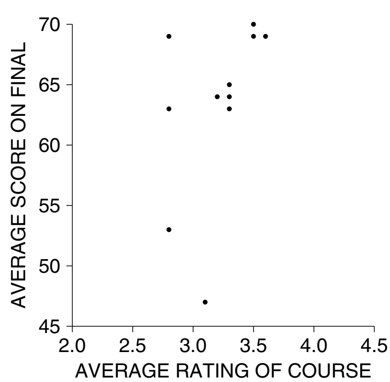
2. (a) False. The scatter diagram slopes down: if x is below average, y is generally above average.

   (b) False. For example, height at age 9 is usually less than height at age 18, but the correlation is positive.

3. (a) Height at 16 and 18. It is easier to predict 2 years ahead than 14.
   (b) Height. Environment, personality, etc. affect weight more than height, and introduce more variability around the line.
   (c) Age 4: by age 18, these other factors have had more time to introduce variability.
4. Somewhat higher; see exercise set B, pp. 145ff.
5. (a) 7: the line $y = 2x - 1$ goes through (1, 1) and (2, 3), then through (4, 7).
   (b) Not possible: (1, 1), (2, 3), and (3, 4) do not lie on a line.
6. No. Data set (ii) is obtained by adding 3 to the y-values in data set (i), so r has to be the same for both. See p. 141.
7. No: section 4.
8. False. The data are cross-sectional, not longitudinal. Younger people were born later and educational levels have been going up over time.
9. (a) False: $r \approx -0.57$. The two sections (C and I) that liked their TA best did worst.
   (b) True: there is no pattern in the scatter diagram ($r \approx 0.12$).
   (c) False: there is a moderate, positive association ($r \approx 0.46$).







10. (a) True. The test-takers are a self-selected group, and good students are more likely to take the test. Hence, if a larger percentage of high school graduates take the test, the average skill level goes down.
    (b) False. The average in Connecticut is lower—but the reason may be that in Connecticut, a higher percentage of students take the test.

    Comment. See note 14 to chapter 9.

11. Quite a bit less. See section 4, and exercise set B on pp. 145ff.

12. (a) Education is measured in whole years (years of schooling completed).
    (b) Some dots correspond to many couples. For example, in many families, the husband and the wife both have 12 years of education (high school degree). All such families are plotted on the (12, 12) dot.
    (c) A (iv) B (iii) C (i)

Chapter 10. Regression

1. A (i) B (iii) C (ii)

2. (a) 112. Work: 115 is 1 SD above average at age 18, so the estimate at age 35 is above average by r SDs. This is $0.8 \times 15 = 12$ points.
   (b) 112. See section 3.

3. (a) 64 inches (b) 62 inches (c) 63 inches, the average (d) same as (c)

   Work for (a). The husband is 4 inches, or $4/2.7 \approx 1.5$ SDs above average in height. The wife is predicted to be above average in height by

   $$r \times 1.5 = 0.25 \times 1.5 \approx 0.4 \text{ SDs}.$$

   This is $0.4 \times 2.5 = 1$ inch.

4. (a) 15 years.
   (b) 13.5 years.
   (c) Appearances are deceiving; all that is going on here is the regression effect (section 5).

5. (a) False (unless there is something special about the SDs and the point of averages).
   (b) False: r measures association not causation.
   (c) True.
   (d) True: the correlation between y and x equals the correlation between x and y.
   (e) False: r measures association not causation.

6. Dashed: y on x.

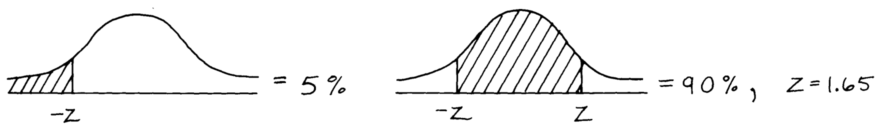
   Dotted: SD line.

   Solid: x on y.

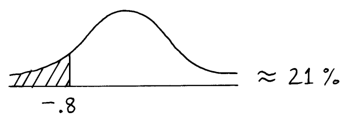
   See section 5.

7. So far, it looks like the regression effect (section 4).
8. The regression effect can't explain a change in the average for the whole population. The data suggest that patients are more relaxed the second time.
9. (a) 21% (b) 65% (c) 50% (d) 50%

   Work for (a). This student was 1.65 SDs below average on the midterm—



   He should be $r \times 1.65 = 0.5 \times 1.65 \approx 0.8$ SDs below average on the final. Now for the percentile rank on the final—



10. False. This person is likely to be between the 40th and 50th percentiles—regression effect (sections 3–4).

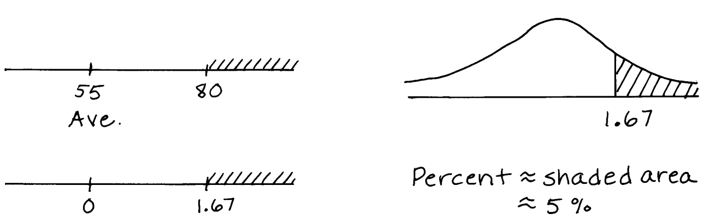
Chapter 11. The R.M.S. Error for Regression

1. Option (v): see p. 186.
2. Something is wrong. GPAs run from 0 to 4. If you predict 2, the maximum error is 2. The computer should be doing better than that.
3. (a) $\sqrt{1 - 0.8^2} \times 2.5 = 1.5$ inches.

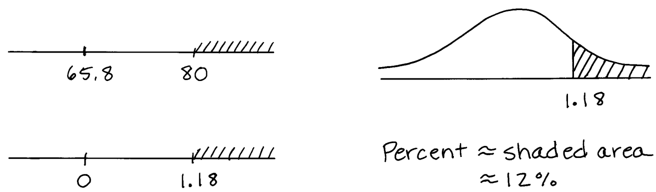
   (b) $\sqrt{1 - 0.8^2} \times 1.7 \approx 1.0$ inches.
4. (a) r.m.s. error $= \sqrt{1 - r^2} \times$ SD of final scores $= 12$.

   (b) 65.8. This student is 30/25 = 1.2 SDs above average on the midterm, and should be above average on the final by $r \times 1.2 = 0.72$ SDs. That is 10.8 points.

   (c) 12, the r.m.s. error: see part (a). It is OK to use the r.m.s. error inside a strip because the diagram is football-shaped.
5. (a) The answer is about 5%.



(b) New average $\approx$ 65.8, new SD = 12, $(80 - 65.8)/12 \approx 1.18$; the answer is about 12%.



6. No. The conclusion seems right, but does not follow from the data. It could be, for example, that better students spend more time doing homework anyway.
7. Option (iii) is right: regression effect.
8. (a) The procedure is subject to error; replication gives more accuracy.
   (b) Two examples of replication: (i) basing course grades on a midterm and final, not just the final; (ii) getting a second opinion before surgery.
9. This is probably the regression effect.
10. If the regression method was used, and you plot the "2006 predicted" against the "2005 actual," you have to get a line—the regression line. Here, the plot is not linear. The regression method was not used to predict 2006 from 2005. (The "actual 2006" column is not needed for the solution.)
11. There are a lot of people with 9, 13, and 17 years of education; very few with 8, 12 or 16: and nobody with 0, although three people have 1 year of education. It looks as if 1 year got added to educational levels, by mistake.

    Comment. The Current Population Survey has public use data sets, which have a fairly complicated file structure. A programmer made an error reading the file, and we caught it by looking at this kind of picture.
12. True, although the difference is pretty small. The correlation is negative: the men with 20 years of eduction should have blood pressures which are, on average, below the grand average by about 3 mm (by the regression method). In other words, the men with 20 years of education should have average blood pressures of about 116 mm.

Chapter 12. The Regression Line

1. In a run of 1 SD, the regression line rises $r \times$SD. The slope is $0.60 \times 20/10 = 1.2$ final points per midterm point. The intercept is $55 - 1.2 \times 70 = -29$. So the equation is

   $$\text{predicted final score} = 1.2 \times \text{midterm score} - 29.$$
2. Predicted income is (\$1600 per inch) $\times$ height $-$ \$81,400. Taller people make more money, on average. Probably, this reflects other variables in family background; although looking every inch an executive may not hurt.

3. Obviously not. The slope means that the 151-pound men are taller, on average, than the 150-pound men—by around 0.0267 inches. Similarly, the men who weigh 152 pounds are on average a little taller than the men who weigh 151 pounds. And so forth.

4. (a) The r.m.s. error is about 1. The points are 1 or so above or below the line.
   (b) No. The scatter diagram slopes down, and so would the regression line; this line is horizontal.

5. This is not legitimate. There are two regression lines. The one for predicting husbands from wives has slope 0.375. See section 10.5.

6. The slope stays the same, and the intercept goes up by 10%.

7. The slope has to be 0.5, because the regression line has to go through the point of averages.

8. (a) True. (b) False. (c) False. (d) True. (e) True.
   Comment. See section 12.2.

9. This is the regression line of IQ on parental income—in disguise (section 10.2). The slope is

   $r \times$ SD of IQ/SD of income $= 0.50 \times 15/45{,}000 = 1/6000$.

10. \$150,000 is likely to be too high. You need the other regression line, for income on IQ (section 10.5). A better estimate is \$105,000.

11. Something is wrong with the equation. The regression line has to go through the point of averages. This line doesn't.

12. No. There doesn't even seem to be any consistent direction to the association.

Part IV. Probability

Chapter 13. What Are The Chances?

1. (a) False. Chances have to be between 0% (can't happen) and 100% (must happen). See p. 223.
   (b) True. The event will happen about 9 times out of 10, and the opposite event will happen the remaining 1 time out of 10. See p. 223.

2. Option (i) is better: you only have to jump one hurdle rather than two.

3. Option (ii) is better, because there are a lot more ways to win:

   clubs diamonds hearts spades
   hearts diamonds clubs spades
   etc.

4. $4/52 \times 3/51 \times 2/50 \times 1/49 \times 4/48 \approx 1/3{,}000{,}000$.

5. Yes. If the ticket is white, then there are two chances in three to get the number 1, and one chance in three to get 8. And the same for black.

6. (a) True.
   (b) True. See example 2 on p. 226.
   (c) False. The two events are dependent. See example 4 on p. 229. The chance is $1/52 \times 1/51$, not $1/52 \times 1/52$.

7. (a) False. (b) False. (c) True.

   Both sequences are equally likely, having chance $1/2^6$. Of course, there are a lot more sequences with 3 heads and 3 tails, but that's not the question.

8. (a) On one roll, the chance of getting 3 or more spots is $4/6 \approx 0.67$; the chance of getting 3 or more spots on 4 rolls is $(0.67)^4 \approx 0.20$, or 20%.
   (b) None of the rolls show 3 or more spots if all show 2 or fewer, and the chance is $(2/6)^4 \approx 0.012$, or 1.2%
   (c) $100\% - 20\% = 80\%$.

   Comment. Compare the two propositions,

   (i) "not all the rolls show 3 or more spots," and

   (ii) "none of the rolls show 3 or more spots."

   Proposition (i) is easier, and has a bigger chance. For example, suppose the dice land 4 6 5 2. Then (i) holds—the last roll was less than 3. But (ii) does not—the first roll is bigger than 3. (So were the second and third rolls.)

9. (a) $(1/6)^{10} = 1/60{,}466{,}176$.
   (b) $1 - 1/60{,}466{,}176 \approx 1$.
   (c) $(5/6)^{10} \approx 0.16$, or 16%.

   Comment. Not getting 10 sixes is easier than getting 10 non-sixes.

10. Option (ii) is better. You have the same 50–50 chance of winning \$1, but you can't lose.

11. Yes. The tickets are 1 1, 1 2, 1 2, 1 3, 3 1, 3 2, 3 2, 3 3.

12. $3/100 \times 2/99 \times 1/98 \approx 6/1{,}000{,}000$.

Chapter 14. More About Chance

1. (a) The chance is $1/6 \times 1/6 = 1/36$.
   (b) The chance is $6/36 = 1/6$; see figure 1 in chapter 14. Or use part (a) and the addition rule. Here is yet another argument. Imagine one of the dice is white and the other is black: no matter how the white one lands, the black one has 1 chance in 6 to match it.

2. From figure 1 in chapter 14, the chance is 2/36.

3. (a) False. These events aren't mutually exclusive, so you can't add the chances. (To find the chance, read section 14.4.)
   (b) False. Same reason.

4. Option (i) is better. Even if you miss the first time, you get a second try at the money.

5. (a) False. A and B can happen together, with chance $1/3 \times 1/10$. So they aren't mutually exclusive.
   (b) True. If A happens, the chance of B drops to 0. That's an extreme form of dependence.

6. If you want to find the chance that <u>at least one of the two events</u> will happen, check to see if they are <u>mutually exclusive</u>; if so, you can <u>add</u> the chances.

   If you want to find the chance that <u>both events</u> will happen, check to see if they are <u>independent</u>; if so, you can <u>multiply</u> the chances.

7. This is like Chevalier de Méré: the chance is $1 - (3/5)^4 \approx 87\%$.

8. 100%—you can't avoid it.

9. Draw a picture like figure 1—

1 1	1 2	1 3	1 4
2 1	2 2	2 3	2 4
3 1	3 2	3 3	3 4

   For instance, 3 2 means you got 3 from box A and 2 from box B. There are 12 outcomes, and each has chance 1/12.

   (a) The good outcomes are 2 1, 3 1, 3 2; the chance is $3/12 = 25\%$.
   (b) $3/12 = 25\%$.
   (c) $100\% - (25\% + 25\%) = 50\%$.

10. Option (ii) is better. There are 60 rolls, and 60 draws. On each roll you have 2 chances in 6 to win \$1 and 4 chances in 6 to get nothing; but on each draw, you have 3 chances in 6 to win \$1 and 3 chances in 6 to win nothing.

11. (a) $13/52 \times 12/51 \times 11/50 \approx 1\%$.
    (b) $39/52 \times 38/51 \times 37/50 \approx 41\%$.
    (c) The chance of getting all diamonds is 1%, see (a). The chance of not getting all diamonds is 99%.

    Comment. "No diamonds" and "not all diamonds" are two different propositions.

12. Both statements are true. Getting ten heads in a row is very unlikely—before you start tossing. If you get nine heads in a row, however, the chance is 50–50 to get a tenth head. (It is the multiplication of all those 50–50 chances that makes ten heads in a row so unlikely.)

13. By trial and error, $(0.98)^{34} \approx 0.503$, and $(0.98)^{35} \approx 0.493$. With 34 draws, there is a 49.7% chance of getting a red marble; with 35 draws, the chance is 50.7%. The answer is 35. (Logs are a more sophisticated option.)

14. The chances are the same. Each ticket has the same (tiny) chance of winning. If you win the grand prize with one ticket, you can't win the grand prize with the other. So, buying two different tickets doubles your chance—addition rule.

Chapter 15. The Binomial Formula

1. The chance is
$$\frac{6!}{1!5!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^5 \approx 40\%.$$

2. The chance of not getting a six in one roll is 5/6, and option (iii) does it by the multiplication rule. (Or use the binomial formula.)

3. The chance of getting 4 girls is $(1/2)^4 = 1/16$. The chance of getting 3 girls is 4/16, by the binomial formula. The total chance is 5/16, by the addition rule.

4. False. The binomial formula does not apply. The draws are dependent, because they are made without replacement.

5. True. The first person gets $8!/(2!6!) = 28$ committees. (Write the 8 names out in a row; then put down 2 C's and 6 N's under the names, where C means "in the committee" and N means "not in the committee"; the binomial coefficient tells you how many ways there are to do that.)

   By similar reasoning, the second person gets $8!/(5!3!) = 56$ committees.

6. False: $8!/(2!6!) = 8!/(6!2!)$.

7. Yes. There are 10 ways to draw two R's and three G's. Each way has the same chance. The total chance is the sum of these 10 chances, that is, 10 times the common value $(1/10)^2(9/10)^3$. The addition rule applies because the ways are mutually exclusive: for example, if you get R R G G G, you cannot get G G G R R.

8. The chance of getting 2 heads among the first 5 tosses is
$$\frac{5!}{2!3!}\left(\frac{1}{2}\right)^5.$$

   The chance of getting 4 heads among the last 5 tosses is
$$\frac{5!}{4!1!}\left(\frac{1}{2}\right)^5.$$

   The first 5 tosses are independent of the last 5. So the answer is
$$\frac{5!}{2!3!}\left(\frac{1}{2}\right)^5 \times \frac{5!}{4!1!}\left(\frac{1}{2}\right)^5 = \frac{50}{1024} \approx 5\%.$$

9. (a) (i): multiplication rule.
   (b) (viii): the chance is 0.
   (c) (iv): addition rule.
   (d) (viii): the events are not mutually exclusive.
   Comment: the chance is $2/52 - 1/52 \times 1/51$
   (e) (vii)
10. Without—then it's bound to happen.
11. The probabilities are 0.8 of 1% for all-cause mortality, 0.2 of 1% for coronary heart disease, and 25% for lung cancer.

    For part (d), the difference in death rates would be very hard to explain by genetics: these twins are genetically identical. For all-cause mortality or heart disease, the difference would be hard to explain by chance. For lung cancer, the difference can be explained by chance. In all cases, the difference can be explained by the health effects of smoking.

    The arithmetic for (a):

$$\binom{22}{17} = 26{,}334 \qquad \binom{22}{20} = 231$$
$$\binom{22}{18} = 7{,}315 \qquad \binom{22}{21} = 22$$
$$\binom{22}{19} = 1{,}540 \qquad \binom{22}{22} = 1$$

    Next, $2^{22} = 4{,}194{,}304$. So the chance is

$$\frac{26{,}334 + 7{,}315 + 1{,}540 + 231 + 22 + 1}{4{,}194{,}304} \approx 0.8 \text{ of } 1\%.$$

    Comments. The constitutional hypothesis is not tenable for all-cause mortality or death from coronary heart disease. On lung cancer, these data do not refute Fisher. Confounding (e.g., by socio-economic status) is a possibility, but not likely: twins are very similar in many respects.

Chapter 15. Special Review Exercise

1. False. The population got bigger too. You need to look at the number of murders relative to total population size. The population in 1990 was about 249 million, and in 1970 it was about 203 million: 20,273 out of 249 million is actually a bit less than 16,848 out of 203 million.

   Comment. The murder rate declined sharply from 1990 to 2000.
2. (a) No. The comparison was with patients treated by other methods in the past.

(b) No. Ullyot was using historical controls, which is not such a good idea—such studies are often biased in favor of treatment. See section 1.3. (Randomized controlled trials were much less positive about the value of coronary bypass surgery.)

3. False. The selection ratio for the pooled data is 68.5%. This is like Simpson's Paradox (pp. 17ff). A majority of the women took the test in 1977, when the pass rates were higher.

   Technical comment. In the pooled data, the women's pass rate is a weighted average of their pass rates for 1975 and 1977, and must be intermediate between those two values. Likewise for the men. The selection ratio, however, is a more complicated statistic. The selection ratio in the pooled data can be written as

$$\frac{79\times\frac{10}{79}+102\times\frac{18}{102}}{79+102}\Bigg/\frac{1312\times\frac{250}{1312}+1259\times\frac{331}{1259}}{1312+1259}$$

   which equals

$$\frac{79\times 0.1266+102\times 0.1765}{79+102}\Bigg/\frac{1312\times 0.1905+1259\times 0.2629}{1312+1259}.$$

   The women are in the numerator; the men, in the denominator. The weights for the women and for the men differ. The women get more weight on the higher pass rate (102 on 0.1765 vs 79 on 0.1266), while the men get more weight on the lower pass rate (1312 on 0.1905 vs 1259 on 0.2629). If the weights are not constrained, the lower bound on the selection ratio in the pooled data is $0.1266/0.2629 \approx 48.16\%$, the numerator being the smaller of the women's two pass rates and the denominator being the larger of the men's two rates. Likewise, the upper bound is $0.1765/0.1905 \approx 92.65\%$. The observed selection ratio of 68.5% is well inside the permissible range.

4. Only (a) is right; (b) has area about 200%, and (c) has the wrong units.

5. About 1 year. One month is way too small (for instance, 68% of the students can hardly be within one month of average); 5 years is too large.

6. False. These data are cross sectional, not longitudinal. The 50-year-old men were born in 1955, and went to school when educational levels were lower than they are now. Furthermore, the material taught was less relevant to today's economy. In these respects, the 60-year-olds are in worse shape. Education is one reason why the curve of age-specific average incomes flattens out. The effect of education is confounded with the effect of age.

   Comment. Other data suggests that for each birth cohort, average income increases fairly steadily with age.

7. False. From left to right, the areas of the blocks are 10%, 10%, 10%, 30%, 40%. The 30th percentile is 60, and the 60th percentile is 80. Generally, the 60th percentile will not be twice the 30th percentile. (Remember, a percentile is a number on the horizontal axis of the histogram.)

8. (a) True. List (ii) is obtained by doubling each entry on list (i), then adding 1. This change of scale cancels on conversion to standard units (section 5.6).

   (b) False. List (ii) is obtained by doubling each entry on list (i), changing the sign, then adding 1. This changes the signs in standard units.

9. (a) The 70th percentile is about 0.52 in standard units, or 60 in original units. The 80th percentile is about 0.84 in standard units, or 67 in original units. The brothers are separated by about 7 points.

   (b) The sisters are separated by about 9 points.

   Comment. As you go further into the tails of the normal distribution, each extra percentile covers a greater and greater distance in the original units—because the normal curve gets lower and lower.

10. It's fine. (The data are from the March 2005 Current Population Survey.)

11. (a) $r \approx 0.82$.

    (b) Can't be done, the points (8, 9) and (8, 13) cannot fall on the same line. (If all the points fall on a vertical line, the correlation is undefined—not the case here.)

12. (a) True.

    (b) True. The correlation between y and x equals the correlation between x and y.

    (c) False, as (a) and (b) should demonstrate.

    (d) False.

13. (i) Nearly 0. (ii) Nearly 1. (iii) Nearly -1.

14. Solid: (i)

    Dotted: (iii)

    Dashed: (ii)

15. The r.m.s. error of the regression line is $\sqrt{1 - 0.55^2} \times 12 \approx 10$. About 68% of the data is within ± 10 points of the regression line, and 16% is above the line by 10 points or more. (The other 16% is below the line by 10 points or more.) The answer is 16%.

16. Statistics—regression effect.

17. The 25th percentile on the final is about 0.675 in standard units, about 73.5 in original units. In other words, of the students who scored around 50 on the midterm, you need to find the percentage scoring over 73.5 on the final. The new average is 60, the new SD is 16, and 73.5 is 0.84 in standard units. The answer is given by the area under the normal curve to the right of 0.84, which is about 20%.

18. There are 30 possible outcomes, all equally likely, shown in the table below.

1 1	1 2	1 3	1 4	1 5	1 6
2 1	2 2	2 3	2 4	2 5	2 6
3 1	3 2	3 3	3 4	3 5	3 6
4 1	4 2	4 3	4 4	4 5	4 6
5 1	5 2	5 3	5 4	5 5	5 6

For instance, 3 2 means you got 3 from box A and 2 from box B.

(a) There are two good outcomes (2 5 and 5 2), the chance is 2/30.

(b) 5/30.

(c) 10/30. You have to look at the table, and count. The good outcomes are shown below, the bad ones are replaced by "x x."

x x	x x	1 3	1 4	1 5	1 6
x x	x x	x x	x x	2 5	2 6
3 1	x x	x x	x x	x x	x x
4 1	x x	x x	x x	x x	x x
5 1	5 2	x x	x x	x x	x x

19. (a) You have to get 3 non-hearts (clubs, diamonds, spades), then a heart. The chance is $39/52 \times 38/51 \times 37/50 \times 13/49 \approx 11\%$.

(b) $13/49 \approx 27\%$. Here, you are asked to compute a conditional chance—of getting a heart on the 4th card, given the first 3 cards were not hearts.

20. If you get 7 heads, you automatically get 3 tails. Use the binomial formula (chapter 15):

$$\frac{10!}{7!3!}\left(\frac{1}{2}\right)^{10} = \frac{120}{1024} \approx 12\%.$$

PART V. Chance Variability

Chapter 16. The Law of Averages

1. Option (iii) is the best. The chance error is not likely to be 0 exactly, but should be small relative to the number of draws.

2. Now option (i) is it.

3. Both are wrong. Luck and the law of averages have nothing to do with it—the chances stay the same, every time.

4. (a) 60 rolls. With more rolls, the percentage of aces will be closer to $16\frac{2}{3}\%$. You want the percentage of aces to be far from $16\frac{2}{3}\%$. Chance error in the percentages is working against you, choose the smaller number of rolls.

(b) 600 rolls. Now chance error in the percentages is working for you, choose the larger number of rolls.

(c) 600 rolls. Like (b).

(d) 60 rolls. As you roll more and more, there get to be more and more possibilities, no particular one can be very likely. Take a more extreme case: with 6000 rolls, you can get 1000 aces, or 1001, or 1002, ... There are lots of possibilities, each one individually has a small chance.

5. False. If the number of heads is 50, the percentage of heads is 50%. (It is not so likely that the number of heads will be 50 exactly, section 1.)

6. Possibility (i) is better. Reason: 10/15 = 20/30 = 2/3. Option (i) is like tossing a coin 15 times, and asking for 2/3 or more heads. Option (ii) ups the number of tosses to 30. With the bigger number of tosses, you are less likely to get 2/3 or more heads. This is like exercise 4(a).

7. The score is like the sum of 25 draws from the box $\boxed{4\ -1\ -1\ -1\ -1}$.

8. The net gain is like the sum of 50 draws from a box with 4 tickets marked "\$8" and 34 tickets marked "−\$1."

9. Choose (ii). With more draws, the percentage of reds is likely to be closer to the percentage in the box, therefore, above 50%. (When the percentage of reds among the draws is above 50%, more reds are drawn than blues, and you win the dollar.) Exercise 9 is like exercise 4(b) or 6.

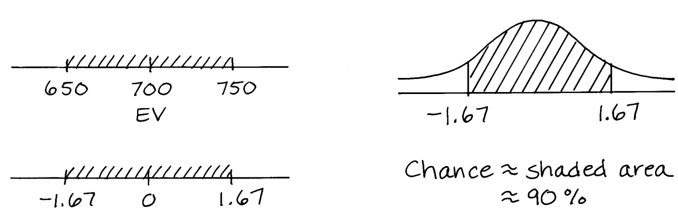
10. (a) 30/200 = 0.15. (b) −0.1. (c) Average = sum/200.

(d) The same: 5/200 = 0.025, so the options describe the same event in different language.

Chapter 17. The Expected Value and Standard Error

1. (a) 100, 1000.

(b) The average of the box is 7 and the SD is 3. So the expected value for the sum is $100 \times 7 = 700$ and the SE is $\sqrt{100} \times 3 = 30$. The sum will be around 700, give or take 30 or so. The chance is about 90%.



2. (a) The net gain is like the sum of 100 draws made at random with replacement from a box with 12 tickets marked "\$2" and 26 tickets marked "−\$1." The average of the box is $-\$2/38 \approx -\$0.05$, and the SD is

$$[\$2 - (-\$1)] \times \sqrt{\frac{12}{38} \times \frac{26}{38}} \approx \$1.39$$

The net gain will be around $100 \times (-\$0.05) = -\$5$, give or take

$$\sqrt{100} \times \$1.39 = \$14 \text{ or so.}$$

(b) The number of wins is like the sum of 100 draws made from a box with 12 tickets marked "1" and 26 marked "0." The number will be around 32, give or take 5 or so.

(c) Both bets pay 2 to 1, but roulette gives you a better chance of winning—12/38 ≈ 32% compared to 25% for Keno. You lose faster at Keno.

3. (a) (iii), see section 4. (b) (i) (c) (v) (d) (iv) (e) (ii)

4. The number of aces in 180 rolls of a die is like the sum of 180 draws from a box with 1 ticket marked "1" and 5 tickets marked "0." The number of aces will be around 30, give or take 5 or so. There is about a 99.7% chance that the number of aces will be in the range 15 to 45. About 99.7% of the people should get a number in that range.

5. The larger number is worse; the chance error is likely to be larger (in absolute terms) with the larger number of throws.

6.

12	chance error in the sum of the draws
45	observed value for the number of 1's
187	observed value for the sum of the draws
25	expected value for the number of 3's
50	expected value for the number of 1's
175	expected value for the sum of the draws
5	standard error for the number of 1's
32	observed value for the number of 3's

7. (a) $321/100 = 3.21$. (b) $3.78 \times 100 = 378$.

(c) The average of the draws will be between 3 and 4 when the sum is between 300 and 400. The expected value for the sum is 350, and the SE is

$$\sqrt{100} \times 1.7 = 17,$$

so the chance is about 99.7%.

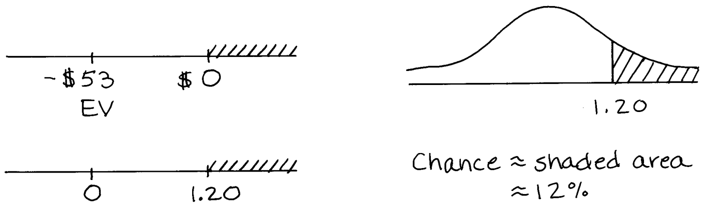
8. (a) Option (ii) is right. The sum is equally likely to go up or down 1 on each draw, just like the difference. Option (i) is out, you can't add words. With option (iii), the sum can't go up. With option (iv), the sum can't go down. With option (v), the sum has a chance to stand still, but the difference has to go up or down.

(b) Expected value $= 0$, SE $= \sqrt{100} = 10$.

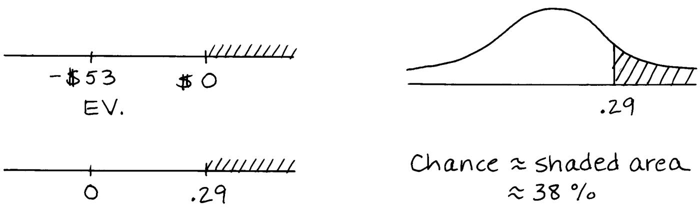
9. (a) is false, (b) and (c) are true. The reason: if you play (i), the net gain is like the sum of 1000 draws from a box with 12 tickets marked \$2 and 26 marked $-\$1$.

The average of the the box is $-\$2/38$ and the SD is about \$1.39. The expected value for the net gain is $-\$53$ and the SE is \$44. If you play (ii), there is another box, with a bigger SD. The net gain has the same expected value but the SE is \$182. Chance variability helps you overcome the negative expected value, so you are more likely to come out ahead with B. Chance variability also makes it more likely that you will lose big.

The chance of coming out ahead with (i) is about 12%.



The chance of coming out ahead with (ii) is about 38%.



10. Statement (iii) is false. The SE goes not go up by the full factor of 2, but only $\sqrt{2} \approx 1.4$. Unless there is something rather strange about the box, the chance that the sum is between 700 and 900 will be quite a bit more than 75%.

11. You have to change the box, to $\boxed{0\ 0\ 0\ 1\ 3}$. The average of the box is 0.8, and the SD is about 1.2. So the sum will be around 80, give or take 12 or so.

12. In all three cases, the expected value is 400 and the SE is 20: the expected value and SE are computed from the box, not the draws. The chance errors are 31, -14, and 17. Remember, Sum = Expected Value + Chance Error.

13. The number of A's is like the sum of 1000 draws made at random with replacement from $\boxed{1\ 0\ 1\ 0\ 0\ 1}$. The SD of the box is 0.50. By the square root law, the SE for number of A's is $\sqrt{1000} \times 0.50$. The number of B's is like the sum of draws made at random with replacement from $\boxed{0\ 0\ 0\ 1\ 0\ 0}$. The SD of the box is $\sqrt{1/6 \times 5/6} \approx 0.37$, so the SE for the number of B's is $\sqrt{1000} \times 0.37$. The SE for the number of A's is larger than for B's. The number of A's is more likely to be 10 or more above expected, because the SE is larger.

    Comment for mathematicians. More generally, let N be the number of draws:

a careful asymptotic argument would separate the cases $N = 6n + r$ for $r = 0, 1, 2, 3, 4, 5$. A delicate case is $r = 2$.

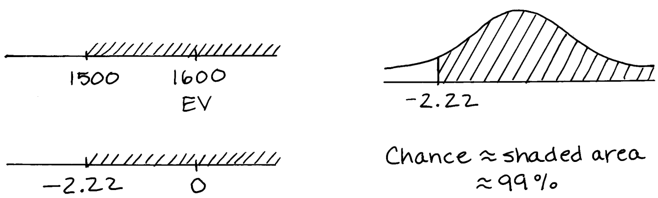
14. The house has 36 chances out of 38 to break even, and 2 chances in 38 to win \$2. In round numbers, the house makes money 5 times, give or take 2 or so. It wins about \$10, give or take \$4 or so.

    Comment. Part (b) follows from (a).

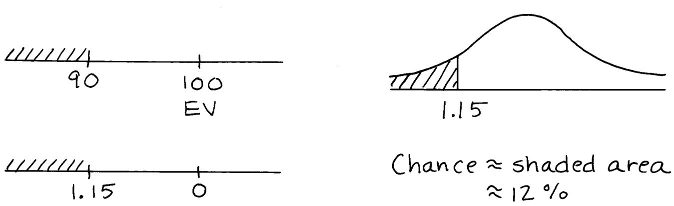
Chapter 18. The Normal Approximation for Probability Histograms

1. 20, 25.

2. (a) The average of the box is 4 and the SD is about 2.24. The EV for the sum is 1600 and the SE is about $\sqrt{400} \times 2.24 \approx 45$. The chance is about 99%.



   (b) The number of 3's is like the sum of 400 draws from the box $\boxed{0\ 1\ 0\ 0}$. The expected number is 100 and the SE is 8.66. The chance is about 12%.



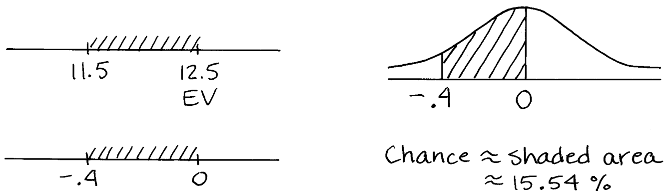
   Comment. For more accuracy, use the continuity correction on part (b),

3. The chance that the sum will be in the interval from 10 to 20 inclusive equals the area under <u>the probability histogram</u> between <u>9.5</u> and <u>20.5</u>. The normal curve is the approximation, the histogram is the truth. You start at 9.5 on the left to catch the block over 10, representing the chance that the sum will be 10. Likewise, you stop at 20.5 on the right to catch the block over 20.

4. If you get 12 heads, you automatically get 13 tails. The chance of getting 12 heads can be found using the method of chapter 15:

$$\binom{25}{12} \times \left(\frac{1}{2}\right)^{25} = \frac{5{,}200{,}300}{33{,}554{,}432} \approx 15.50\%$$

   Or, use the method of example 1 on p. 317. The expected number of heads is 12.5, and the SE is 2.5. The normal curve gives a very good approximation:



5. (i) is the probability histogram for the sum, it's like the normal curve.
(ii) is the probability histogram for the product, it's like figure 10, p. 323.
(iii) is the histogram for the numbers drawn, it's like a histogram for the numbers in the box.

6. Something is wrong with COIN. With a million tosses, the number of heads should be around 500,000 give or take 500 or so. COIN is 4 SEs too high. (The SE for the number of heads is a tiny fraction of the number of tosses.)

7. Option (ii) is right. With (i), all the values have equal chances, and that's wrong (figure 1, p. 311, bottom panel.)

8. (a) is true, (b) is false, (c) is false, (d) is true.
The expected value is computed without error, as 50. The 5 represents the likely size of the chance error in the number of heads, not in the expected value.

9. (a) is true, (b) is false, you don't have enough draws to use the normal approximation. See exercise 6, p. 324.

Technical comment. If you use the continuity correction, the normal approximation is not ridiculous—for this symmetric interval.

10. Both statements are true. With more tosses, the probability histogram gets closer to the normal curve. You get a feeling for how many draws are enough by looking at the pictures in the chapter.

11. The sum of the draws is 270, with EV $= 250$ and SE $= 15$; the chance error is 20, which is 1.33 SEs. The number of 1's is 17, with EV $= 25$ and SE ≈ 4.33; the chance error is -8, which is about -1.8 SEs. The number of 2's is 54, with EV $= 50$ and SE $= 5$; the chance error is 4, which is 0.8 SEs.
(a) number of 2's (b) sum of the draws

12. (a) Can't be done. For example, the box could have 4 tickets marked "1" and 6 marked "−10," or 4 marked "3" and 6 marked "−2." With these two boxes, the chances would be quite different.
(b) Can be done, using the normal approximation—that's why the average and SD are so useful.

13. Can't be done. You need to know how many 3's there are in the box.

14. This is like drawing at random with replacement from a box with 4 tickets marked "1" and 6 marked "0," and asking for the chance that the sum of the

draws will be 425 or more. The normal approximation will be fine. (The average and SD of the original numbers in the box are irrelevant.)

15. Very fishy. Very. The chance of getting a red marble is 40%, so the expected number of reds is 40; the chance of getting 39 or fewer is about 50% (or a little less). All 10 counts are below 40—an event of probability $1/2^{10}$.

Part VI. Sampling

Chapter 19. Sample Surveys

1. False. Simple random samples are impractical for this purpose (section 4).
2. No, because of response bias. The subjects could try to give the interviewers the pleasing answer, rather than the true answer. (In this study, many respondents said they were using the product when they really weren't.)
3. No. This county might be exceptional. (It was.)
4. The people with college degrees were living in more suburban neighborhoods. This was not a good way to draw a sample.
5. This estimate is likely to be too high. With smaller households, the interviewer is less likely to find someone at home. So the survey procedure is, on average, replacing smaller households by larger ones.
6. No. Different sorts of students are more likely or less likely to walk through different parts of the campus at different times, and the chances are difficult or impossible to figure (p. 341).
7. Option (ii) is better. You win if number of heads is between 480 and 520. That's a broader range than (i).
8. Cannot be determined from the information given. The rectangle over a number might represent chance, or it might represent an empirical frequency. You need to know how the histogram was computed.
9. The smaller one. As the size of the hospital goes up, the percentage of male births gets closer to 52%, and is less likely to exceed 55%. See section 16.1.
10. Look at figure 3, chapter 18. The likeliest number of heads is 50: pick that first. Your next two picks should be 49 and 51. Then 48 and 52. And so forth. You should pick 45 through 55. And your chance of winning is about 73%—example 1(b) on p. 317.
11. (a) 10 through 20. (b) 40 through 50.
    (c) To get the center of the range, take 50 from the sum; and then go 5 either way.
    (d) 73%, as in the previous exercise.

12. (a) Only the "high achievers" are listed in the sampling frame.
(b) There seems to be quite a lot of room for bias here. How were the 5,000 students selected? What about non-response bias?

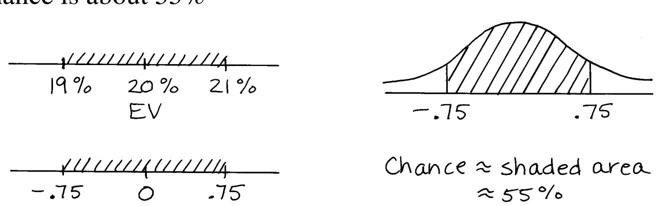
Chapter 20. Chance Errors in Sampling

1.

	No. of heads		Pct. of heads	
Number of tosses	Expected value	SE	Expected value	SE
100	50	5	50%	5%
2,500	1,250	25	50%	1%
10,000	5,000	50	50%	0.5 of 1%
1,000,000	500,000	500	50%	0.05 of 1%

2. The first step is setting up the box model. The number of aces is like the sum of 1000 draws from the box $\boxed{0\ 0\ 0\ 0\ 0\ 1}$. The average of the box is 1/6 and the SD is $\sqrt{1/6 \times 5/6} \approx 0.373$. The EV is $1000 \times 1/6 \approx 166.7$. The SE is $\sqrt{1000} \times 0.373 \approx 11.8$. The number of aces will be around 167, give or take 12 or so. The SE for the percentage is 12 out of 1000, which is 1.2%. The percentage of aces should be around $16\frac{2}{3}$%, give or take 1.2% or so.

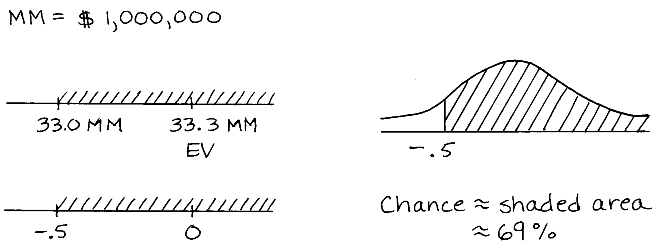
3. (a) There should be 50,000 tickets in the box. The box is the population—all 50,000 forms.
(b) Each ticket shows a 0 (gross income under \$50,000) or a 1 (gross income over \$50,000).
(c) False: the SD of the box is $\sqrt{0.2 \times 0.8} = 0.4$.
(d) True. The draws are the sample.
(e) The number of sample forms with gross incomes over \$50,000 is like the sum of 900 draws from the box. The expected value for the sum is 180. The SE for the sum is $\sqrt{900} \times 0.4 = 12$. The number of sample forms with gross incomes over \$50,000 will be around 180, give or take 12 or so. Now 12 out of 900 is about 1.33%. The percentage of forms in the sample with gross incomes over \$50,000 will be 20.00%, give or take 1.33% or so. The chance is about 55%—



(f) Can't be done with the information given. You need to know the percentage of forms with gross incomes over \$75,000. And you can't use the normal curve, because these data are far from normal.

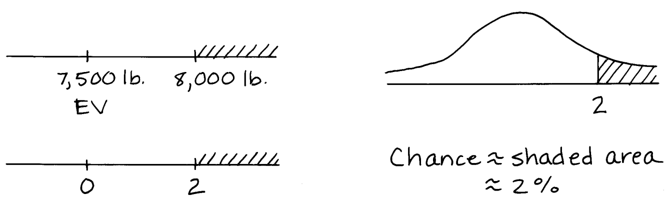
4. (a) 50,000. (b) A gross income. (c) True. (d) True.

   The total gross income is like the sum of 900 draws from the box, whose average is \$37,000. The SD is \$20,000. The expected value for the total is $900 \times \$37{,}000 = \$33{,}300{,}000$. The SE is $\sqrt{900} \times \$20{,}000 = \$600{,}000$. The chance is about 69%.



   In exercise 3, you were classifying and counting, so you had to change the box; here, you are working with a sum, so it would be a bad idea to change the box.

5. The total weight of the group is like the sum of 50 draws from a box. The average of the box is 150 pounds and the SD is 35 pounds. The expected value for the sum is 7,500 pounds and the SE is $\sqrt{50} \times 35 \approx 250$ pounds. The chance is about 2%—maybe a better elevator is needed.



6. Option (ii) is right. In absolute terms, the California sample will be much bigger, and therefore more accurate for estimating percents; the size of the sample relative to the size of the state is basically irrelevant (section 4).

7. (a) is true and (b) is false. The expected value is computed with no chance error: the EV equals the percentage of 1's in the box.

   (c) is true and (d) is false. The percentage of 1's among the draws will be off its expected value, due to chance error. The SE for the number of 1's among the draws is $\sqrt{500} \times \sqrt{0.25 \times 0.75} \approx 9.7$. The SE for the percentage of 1's among the draws is $(9.7/500) \times 100\% \approx 2\%$.

   (e) is true and (f) is false: you know what's in the box.

8. The expected value for the percentage of democrats in the sample is 40%. The normal curve can be used to figure chances here (p. 364), because the sample is reasonably large. Half the area under the curve is to the right of 0, and 0 in standard units is just the expected value. In other words, there is just about

a 50% chance for the sample percentage to exceed the population percentage. So, 40% goes into the blank.

9. The number of 1's is the sum of the draws. The expected value is 200, and the SE is about 12.

10. It's fine. (This chapter is about the SE for percentages, but the idea of computing SEs for numbers should not disappear.)

11. (a) 357, 340. (b) 71.4%, 68%.

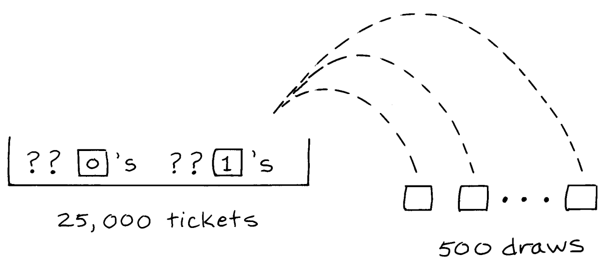
    Comment. The expected value is computed from the box; the observed value, from the sample.

12. The total number of interviews is like the sum of 400 draws from a box. The average of the box is 2.38, and the SD is 1.87. The total number of interviews will be around $400 \times 2.38 = 952$, give or take $\sqrt{400} \times 1.87 \approx 37$ or so.

Chapter 21. The Accuracy of Percentages

1. The number in the sample who approve of the Mayor is like the sum of 1000 draws made at random without replacement from a box with a ticket for each registered voter in the town; the ticket is marked "1" if the voter approves of the Mayor and "0" otherwise.

2. (a) The sample is like 500 draws from a box with 25,000 tickets; each ticket is marked 1 (has a computer) or 0 (does not have a computer). The number of sample households with computers is like the sum of the draws.



   The fraction of 1's in the box is unknown, but can be estimated by the fraction in the sample, as 239/500 = 0.478. On this basis, the SD of the box is estimated as $\sqrt{0.478 \times 0.522} \approx 0.50$. The SE for the number of sample households with computers is estimated as $\sqrt{500} \times 0.50 \approx 11$, and 11 out of 500 is 2.2%. The percentage of households in the town with internet access is estimated as 47.8%, and the estimate is likely to be off by 2.2% or so.

   (b) The 95%-confidence interval is 47.8% ± 4.4%.

   Comment. You need to estimate the percentage for the town from data for the town: the national figures may not apply. (In this case, the town seems pretty close to the national average.)

3. (a) 1.4%, 0.5 of 1%.

(b) Hard to do: the box is lopsided and the normal approximation won't work very well. (See exercises 5–6 on p. 324; exercises 3–4 on p. 383.)

4. In the sample, $(172 + 207)/500 = 379/500 = 75.8\%$ of the households owned cars. The percentage of households in the town with cars is estimated as 75.8%, give or take 1.9% or so.

5. (a) The box has millions of tickets, one for each 17-year-old in school that year. Tickets are marked 1 for those who knew that Chaucer wrote *The Canterbury Tales*, and 0 for the others. The data are like 6000 draws from the box, and the number of students in the sample who know the answer is like the sum of the draws. The fraction of 1's in the box can be estimated from the sample as 0.361. On this basis, the SD of the box is estimated as $\sqrt{0.361 \times 0.639} \approx 0.48$. The SE for the number of students in the sample who know the answer is estimated as $\sqrt{6000} \times 0.48 \approx 37$. The SE for the percentage is 37/6000, which is about 0.6 of 1%. The percentage of students in the population who know the answer is estimated as 36.1%, give or take 0.6 of 1% or so. The 95%-confidence interval is $36.1\% \pm 1.2\%$.

   (b) $95.2\% \pm 0.6$ of 1%.

6. False, because of chance error. The sample percentage is likely to be close to the population percentage, but not exactly equal. The SE for the percentage says how far off you can expect to be.

7. This is not the right SE. We do not have a simple random sample of 252 days, and the daily changes are dependent: each day's closing price is the next day's opening price.

8. Yes. This is like example 1 on p. 378: the estimate is based on a simple random sample.

9. This is not the right SE. The bank has mixed up 73¢ with 73%. (The right way to figure the SE for an average will be explained in chapter 23.)

10. Your net gain is like the sum of 100 draws from a box with 6 tickets marked "\$11" and 94 marked "−\$1." The average of the box is −\$0.28, and the SD is about \$2.85. Your net gain will be around −\$28, give or take \$28 or so.

11. $710/100 = 7.1$, so the two options describe the same event in different words. They are the same.

12. Option (ii) is it. For example, about 95% of the estimates will be right to within 2 SEs, about 99.7% of them will be right to within 3 SEs, and so forth.

13. (i) is irrelevant, (ii) is a histogram for the numbers drawn, and (iii) is a probability histogram for the sum. Reason: (iii) looks like the normal; (ii) looks like a histogram for the contents of the box; (i) would allow 3 and 4 among the draws.

14. In all three cases, the expected value is 500 and the SE is about 16. You know

what is in the box, the expected value and SE can be computed exactly and do not depend on the data. The three chance errors are 29, −16, and 14.

15. (a,b) Estimated from the data as. Here, you do not know the composition of the box; this must be estimated from the data.

Chapter 22. Measuring Employment and Unemployment

1. (a) False (section 4).
   (b) They were outside the labor force (section 3)—going to school, retired, keeping house, etc.
2. False. This is not a simple random sample (section 5).
3. The estimate is 7.0 million, and the estimated SE is 0.1 million (section 5).
4. Not much (section 7).
5. A simple random sample is drawn at random without replacement (p. 340).
6. Drawing without replacement gives a more accurate estimate for the percentage of 1's in the box, because you are sure to get a new ticket on each draw. (For an extreme example, think about 250 draws: drawing without replacement gives a perfect estimate; not so when drawing with replacement.) Also see p. 368.

   Comment. Drawing with replacement leads to easier formulas for the SE. That is why we study it. The square root law is exact for drawing with replacement, and approximate for drawing without replacement. Exercise 6 is asking about the accuracy of the estimate for the percentage of 1's in the box; it is not asking which formula for the SE is easier or more accurate.
7. (a) This is a probability sample (section 19.4).
   (b) This is not a simple random sample. A simple random sample would give you a chance to draw households from all 5 districts, here you only get 2 out of the 5.

   Comment. There is no selection bias, the procedure is fair and impartial, and in fact, every household has the same chance to get into the sample.
8. (a) This is a probability sample (section 19.4).
   (b) This is not a simple random sample. For example, the auditors can't get items 17 and 18 in the same sample: with a simple random sample, they could.

   Comment. This kind of sample is called a *list sample*, or a *systematic sample*. There is no selection bias, the procedure is fair and impartial, and in fact, every item has the same chance to get into the sample. List samples are easier to draw than simple random samples, and for many purposes they work just fine. You have to be a little careful in figuring the SE, though; and it's better to use several random starts than one.

9. No, because of response bias. The attitude of the interviewers could well affect the responses. (And in this example, it did.)

10. The total of the numbers tossed is like the sum of 200 draws from the box

    $\boxed{1\ 2\ 3\ 4\ 5\ 6}$.

    The average of the box is 3.5, so the total should be around

    $200 \times 3.5 = 700.$

    Boyd is giving us the old razzle dazzle.

11. (a) (i) (b) (ii) (c) (ii) (iii)

    Comment. The expected value of the *sample percentage* equals the *population percentage*. This is so even after the sample is drawn—the expected value is sort of the average over all possible samples, not just the particular sample you happened to draw. In the frequency theory, it is a mistake to say that the expected value of the *population percentage* equals the *sample percentage*. See section 21.3.

12. (a) True.

    (b) True.

    (c) False, the normal curve will not work here, the data are too skewed. (For example, the percentage cannot be negative, but $1.04 - 1.46 < 0$.)

Chapter 23. The Accuracy of Averages

1. The sum of 400 draws will be around $400 \times 100 = 40{,}000$, give or take

    $\sqrt{400} \times 20 = 400$ or so.

    The average of 400 draws will be around 100, give or take 1 or so.

    (a) The chance is almost 100%.

    (b) The range is "expected value $\pm$ 1 SE," so the chance is about 68%.

    Moral: the SE and the SD are different.

2. (a) True. That's what the SE does for you (pp. 417, 422).

    (b) True. The 68%-confidence interval is "sample average $\pm$ SE for average."

    (c) False (p. 417).

3. Model: there is a box with 50,000 tickets, one for each household in the town. The ticket shows the commute distance for the head of household. The data are like 1000 draws from the box. The SD of the box is unknown, but can be estimated by the SD of the data, as 9.0 miles. The SE for the sum of the draws is estimated as $\sqrt{1000} \times 9 \approx 285$ miles, and the SE for the average is estimated as $285/1000 \approx 0.3$ miles.

    (a) The average commute distance of all 50,000 heads of households in the town is estimated as <u>8.7 miles</u>; this estimate is likely to be off by <u>0.3 miles</u> or so.

    (b) 8.7 miles $\pm$ 0.6 miles.

Comment. The normal curve is used to approximate the probability histogram for the sample average, not the histogram for the data; the data are skewed, with a long right hand tail.

4. Can't be done with the information given. This is a simple random sample of households, but a cluster sample of people. The cluster is the household, and people in a household are likely to be similar with respect to commuting. For example, if a household is far from the center of town, all the occupants are likely to have a long commute. The SE is going to be bigger than the SE for a simple random sample of 2500 persons: section 22.5.

5. Model: there is a box with 50,000 tickets, one for each household in the town. The ticket is marked 1 if the head of household commutes by car; otherwise, 0. The data are like 1000 draws from the box. The fraction of 1's in the box is unknown, but can be estimated by the fraction in the sample, as 0.721. On this basis, the SD of the box is estimated as $\sqrt{0.721 \times 0.279} \approx 0.45$. The SE for the number of 1's in 1000 draws is estimated as $\sqrt{1000} \times 0.45 \approx 14$. The SE for the percentage of 1's is 14/1000, or 1.4%. The percentage of 1's in the box is estimated as 72.1%, give or take 1.4% or so. The 95%-confidence interval is 72.1% $\pm$ 2.8%.

   Comment. We have a simple random sample of households, and are making an inference about households.

6. (a) Can't be done, you would need to use the half sample method (section 22.5).
   (b) SE $\approx$ 1.

7. Option (iii) is it, this is a sample of convenience (p. 424).

8. (a) True: the interval is "average $\pm$ SE."
   (b) True: section 21.3.
   (c) The data don't follow the normal curve, but the 68% might be right; you need the data to tell. (To see that the data don't follow the curve: enrollments can't be negative, but the SD was a lot bigger than the average, so there must have been a long right hand tail.)
   (d) False: 325 is not the SD. (The data aren't normal, which is another problem.)
   (e) False. The normal curve is being used on the probability histogram for the sample average, not the data (pp. 411 and 418–19).

9. 1.7 $\pm$ 0.1.

10. (a) True: the SE is estimated from the sample data, as on p. 416.
   (b) False. There is no such thing as a 95%-confidence interval for the *sample average*, you know the sample average. It's the population average that you have to worry about (pp. 385–86).
   (c) True.

(d) False. This confuses the SD with the SE. And it's ridiculous, because a household must have a whole number of persons (1, or 2, or 3, and so forth).

(e) False. For instance, if household size followed the normal curve, there would be many households with a negative number of occupants; we're not ready for that.

(f) True. See pp. 411 and 418–19. Even though household size does not follow the normal curve, you can still use the normal curve to approximate the probability histogram for the sample average.

11. There is too much spread in the histogram: the SE for the average is only about 0.3, there is a lot of area outside the range EV $\pm$ 3 SE.

12. This is not a 95%-confidence interval: the class is a sample of convenience, not a probability sample.

Chapter 23. Special Review Exercises

1. No. The teachers might have been tempted to put the poorer children into the treatment group. (In fact, this seems to have happened; children in the treatment group were significantly smaller in physical size than the controls, which probably biased the study against the treatment.)

2. No. In badly designed studies of surgery, healthier patients are more likely to be picked for the operation, leaving sicker patients to be the controls. Being a control in a badly designed study may be a marker of illness (section 1.2), but it is unlikely to be a cause of illness.

3. (a) Yes. Other things being equal, the drinkers will have higher rates of oral cancer than the non-drinkers, due to the alcohol.
   (b) Option (ii) is right: section 2.5.

4. (a) Yes.
   (b) Yes: we're looking at people who died in 2005. The older ones were born earlier, and are more likely to be right handed.
   (c) The data don't distinguish between (a) and (b), and don't tell you much.

5. Choose (iii). The salary data have a long right hand tail, the average is bigger than the median. The total of a list is

   number of entries $\times$ the average.

   The total cannot be computed from the median.

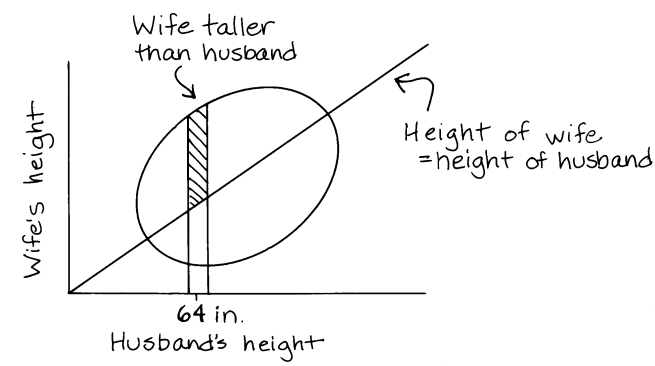
6. 10 years. If the SD were 1 year or 2 years, a lot of people would be more than 5 SDs from average. If the SD were 25 years, hardly anybody would be more than 1 SD from average.

7. Negative: as chlorophyll concentration becomes bigger, the water gets murkier, and Secchi depth becomes smaller.

8. Yes.

9. Guess 570, using the regression method. You have about a 68% chance to be right to within 1 r.m.s. error, which is about 86 points.

10. The scatter diagram is sketched below. The families where the husband is about $5'\,4'' = 64''$ tall are in the vertical strip. For these husbands to be shorter than their wives, the wife must be taller than $64''$. In this strip, the average height of the wives is $62.1''$ (the new average), and the SD is $2.4''$ (the new SD). The wives have to be taller than $64''$, that is, more than 0.8 new SDs above the new average. The percentage is about the area under the normal curve to the right of 0.8, that is, 21%.



11. (a) 0.25. This line is the regression line for predicting wife's income from husband's (section 10.2).
    (b) The estimate is likely to be too high. You need to use the other line, for predicting husband from wife (section 10.5).

12. Something is wrong. Both lines are too shallow to estimate the average of x given y. (The lighter line is the regression line for y on x; the heavier line is the SD line.)

13. The chance of getting neither an ace nor a king is

    $44/52 \times 43/51 \times 42/50 \times 41/49 \times 40/48 \approx 42\%.$

    The answer is $100\% - 42\% = 58\%$.

14. Not possible with information given. You need to know how many people got perfect scores on both midterms.

15. For (a, b, c), the answer is

$$\binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{250}{7776} \approx 3\%$$

16. False. Option (i) has chance $1 - (37/38)^{15} \approx 33\%$, while option (ii) has chance $1 - (37/38)^{30} \approx 55\%$. This is like de Méré (section 14.3).

17. The total is like the sum of 20 draws from the box $\boxed{1\ 0\ 0\ 0\ 0\ 1}$. The average of this new box is 0.33, and the SD is 0.47 by the shortcut (section 17.4). The total will be around $20 \times 0.33 \approx 7$, give or take $\sqrt{20} \times 0.47 \approx 2$ or so.

18. If a student answers at random, the score is like the sum of 50 draws from the box $\boxed{2\ \ -1\ \ -1}$. The average of this box is 0, and the SD is 1.41 by the shortcut (section 17.4). The EV for the sum of the draws is $50 \times 0 = 0$; the SE is $\sqrt{50} \times 1.41 \approx 10$. The score will be around 0, give or take 10 or so.

    (a) A cutoff of 50 is 5 SEs, the chance is zilch.

    (b) A cutoff of 10 is 1 SE, the chance is about 16%.

19. The histogram for the data will get closer and closer to the probability histogram for the number of heads in 100 tosses of a coin (section 18.2). This probability histogram is not the normal curve—although it's close (chapter 18, figure 3). Also see pp. 326–27.

20. (a) The normal approximation will be too low, because the curve is lower than the probability histogram at 90.

    (b) The approximation will be about right, because the highs and lows cancel.

21. The number of heads in each group should be around 50, give or take 5 or so; and the groups are independent. Option (i) is right. With (ii), the number of heads is more like 25. With (iii), there is a strong negative correlation from group to group. And with (iv), the SD is more like 10.

22. The EV is 40% and the SE is 2.2% in both parts of the problem. For instance, the SE for the number of blues is $\sqrt{500} \times \sqrt{0.4 \times 0.6} \approx 11$, so the SE for the percentage of blues among the draws is 11/500 or 2.2%. It is a mistake to use the sample fractions when the contents of the box are given.

    (a) The observed value is 218/500 or 43.6%; the chance error is 3.6%.

    (b) The observed value is 191/500 or 38.2%; the chance error is -1.8%.

23. Option (ii) is right. The top histogram is symmetric, the bottom one is skewed. If Box A and Box B were the same, the two histograms would have to be the same (in standard units). See p. 411, and exercise 8 on p. 414.

24. (a) True (section 18.5).

    (b) False. The histogram for the numbers in the box does not change, and could have any shape—depending on how the box was set up in the first place.

    (c) False. The histogram for the draws gets more and more like the histogram for the numbers in the box. You have to distinguish between (i) the probability histogram for the sum of the draws, and (ii) the histogram for the draws as data.

    (d) False (p. 323).

    (e) True.

25. How did they pick the 5 employees? Quite a lot of bias could come in at this stage.

26. The number of blacks in the sample should be around 26, give or take 4.4 or so. The chance of getting 8 or fewer is nearly 0. These jurors were not chosen at random, whatever the Supreme Court thought at the time.

27. (a) 10.8%.

    (b) 1.4%.

    (c) The range from <u>8.0%</u> to <u>13.6%</u> is a 95% confidence interval for the percentage of independents <u>among all registered voters in Hayward</u>.

28. This is a cluster sample, so you can't compute the SE from the information given (section 22.5).

29. (i) is the histogram for the numbers drawn: it looks like (iii), but is a little off due to chance error.
    (ii) is the probability histogram for the average of the draws; it is starting to look like the normal curve, but is still a little irregular (there are only 20 draws).
    (iii) is the histogram for the contents of the box: there are exactly as many 2's as 4's, and twice as many 1's as 2's.

30. False. Accuracy in estimated percentages depends mainly on the sample size, not the population size (section 20.4).

Part VII. Chance Models

Chapter 24. A Model for Measurement Error

1. Model: each measurement equals the exact elevation, plus a draw from the error box. The tickets in the box average out to 0. Their SD is unknown, but can be estimated by the SD of the data, as 30 inches. The SE for the sum of the 25 measurements is estimated as $\sqrt{25} \times 30 = 150$ inches. The SE for the average is estimated as $150/25 = 6$ inches.

    (a) 81,411 inches, 6 inches.

    (b) True.

    (c) False. You don't need a confidence interval for the *sample average*, you know the sample average. It's the elevation of the mountain that needs to be estimated.

    (d) False. This mixes up SD and SE.

    (e) False, same issue.

    (f) False; see exercise 8 on p. 421.

2. (a) The <u>probability histogram for the average of the draws</u> follows the curve.

    (b) The <u>histogram for the data</u> doesn't follow the curve.

3. 299,774 ± 0.6 km/sec.

4. The SD of the distance measurements.

5. If they got the length wrong (as they must have, at least by a little), there is a bias in the speed measurements.

6. False. The average for 2005 is known, a confidence interval is not needed. At a more basic level, the Gauss model does not apply (example 4 on p. 446).

7. Use the SD of the old data, 18 micrograms (example 5 on p. 451). The 95%-interval is 78.1 ± 5.1 micrograms above 1 kilogram.

8. The data are like 100 draws made at random with replacement from a box. The average of the box is about 58 seconds, and the SD is about 2 seconds. The total execution time is like the sum of 100 draws from the box. This will be about $100 \times 58 = 5800$ seconds, give or take $\sqrt{100} \times 2 = 20$ seconds or so.

9. (a) Model: the weights of the sticks are drawn at random from a box. The average of the box is 4 ounces, and the SD is 0.05 ounces. The weight of a package is like the sum of 4 draws. The expected value is $4 \times 4 = 16.0$ ounces. The SE for the sum is $\sqrt{4} \times 0.05 = 0.1$ ounces. The answer: 16 ounces, 0.1 ounces.

   (b) The total weight of 100 packages is like the sum of 400 draws from the box. The expected value is $400 \times 4 = 1600$ ounces $= 100$ pounds. The SE for the sum is $\sqrt{400} \times 0.05 = 1$ ounce. The total weight will be 100 pounds, give or take 1 ounce or so. The range "100 pounds ± 2 ounces" is "expected value ± 2 SE," so the chance is about 95%.

10. False. You use the curve on the probability histogram for the average, not the histogram for the data (pp. 418–19).

11. This procedure invites bias. It would be better to make the measurements some prespecified number of times, and then take the average. Of course, if something went wrong during the measurement process, it might be okay to exclude the result.

Chapter 25. Chance Models in Genetics

1. *First line*. The yellow parent must be y/y. The green parent must be g/y, else, all progeny would be green.

   Second line. Both parents must be g/y, else no yellow progeny.

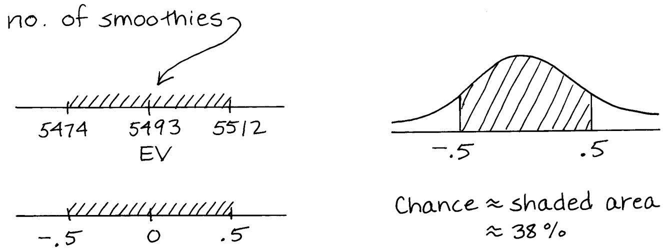
   Third line. Both parents are y/y.

   Fourth line. The yellow parent is y/y. The green parent must be g/g, else there would be yellow progeny.

   Fifth line. One parent must be g/g, else there would be yellow progeny; the other parent can be g/g or g/y.

2. Genetic model: one gene-pair, with two variants, smooth (s) dominant and wrinkled (w) recessive. Crossing s/w with s/w produces smooth with probability 3/4, wrinkled with probability 1/4. Mendel had $5474 + 1850 = 7324$ plants. The number of smoothies is like the sum of 7324 draws from the box $\boxed{1\ 1\ 1\ 0}$.

The expected value is $7324 \times 3/4 = 5493$. He was off by $5493 - 5474 = 19$. The standard error is $\sqrt{7324} \times \sqrt{3/4 \times 1/4} \approx 37$. The chance is about 38%.



3. Genetic model: one gene-pair, with two variants, early (e) and late (l). e/e is early, l/l is late and e/l is intermediate. So intermediate × intermediate should give about 25% early, 50% intermediate, 25% late. With 2500 plants from this cross, the number of intermediates is like the sum of 2500 draws from the box $\boxed{1\ \ 0}$. The expected number is 1250, the SE is 25, and the chance is about 2.5%.

4. (a) No. The man's father contributed the Y-chromosome, and this has nothing to do with baldness.
   (b) Yes. The mother got one X chromosome from her father (the man's maternal grandfather), and may have passed it on to the man.

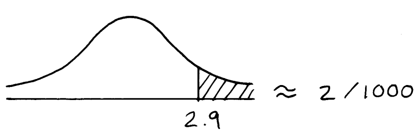
5. (a) No. The child will have at least one A.
   (b) Yes, if both parents are A/a and the child gets a from each parent.
   (c) No. Both parents are a/a, so the child will be a/a too.

Part VIII. Tests of Significance

Chapter 26. Tests of Significance

1. (a) True (p. 479).
   (b) False. The null says it's chance, the alternative says it's real (pp. 477–78).

2. (a) The data are like 3800 draws made at random with replacement from the box $\boxed{\text{?? 0's} \quad \text{?? 1's}}$, with 1 = red.
   (b) Null: the fraction of 1's in the box is 18/38, or 47.4%.
   Alt: the fraction of 1's in the box is more than 18/38.
   (c) The expected number of reds (computed using the null) is 1800. The SD of the box (also computed using the null) is nearly 0.5, so the SE for the number of reds is $\sqrt{3800} \times 0.5 \approx 31$. So $z = (\text{obs} - \text{exp})/\text{SE} = (1890 - 1800)/31 \approx 2.9$, and $P \approx 2/1000$.



(d) Yes.

Comments. (i) This problem is about the number of reds. In the formula for the z-statistic, *obs, exp*, and *SE* all refer to the number of reds. The expected, as always, is computed from the null. In this problem, the null gives the composition of the box, so the SD is computed from the null; it is not estimated from the data (p. 485).

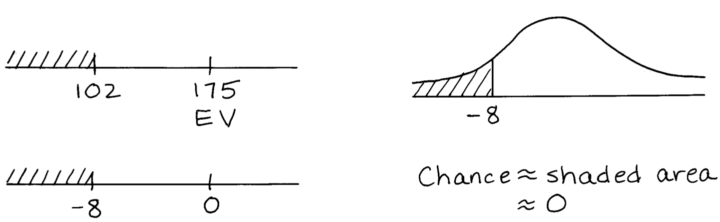
(ii) This problem, and several others below, can be done using one-sided or two-sided tests. The distinction does not matter here; it is discussed in chapter 29.

3. Null hypothesis: the data are like 200 draws from the box $\boxed{1\ 1\ 1\ 0}$, with 1 = blue and 0 = white. The expected number of blues is 150, and the SE is about 6, so z = (obs − exp)/SE = (142 − 150)/6 ≈ -1.3 and P $\approx$ 10%. This is marginal, could be chance.

4. The TA's null: the scores in his section are like 30 draws at random from a box containing all 900 scores. (There is little difference between drawing with or without replacement, because the box is so big.) The null hypothesis specifies the average and the SD of the box, 63 and 20. The EV for the average of the draws is 63, and the SE is 3.65. So z = (obs − exp)/SE = (55 − 63)/3.65 ≈ -2.2, and $P \approx 1\%$. The TA's defense is not good.

5. The box has one ticket for each freshman at the university, showing how many hours per week that student spends at parties. So there are about 3000 tickets in the box. The data are like 100 draws from the box. The null hypothesis says that the average of the box is 7.5 hours. The alternative says that the average is less than 7.5 hours. The observed value for the sample average is 6.6 hours. The SD of the box is not known, but can be estimated from the data as 9 hours. On this basis, the SE for the sample average is estimated as 0.9 hours. Then z = (obs − exp)/SE $\approx$ (6.6 − 7.5)/0.9 = -1. The difference looks like chance.

6. (a) In the best case for Judge Ford, we are tossing a coin 350 times, and asking for the chance of getting 102 heads or fewer. The expected number of heads is 175 and the SE is a little over 9, so $z \approx (102 - 175)/9 \approx -8$. The chance is about 0.



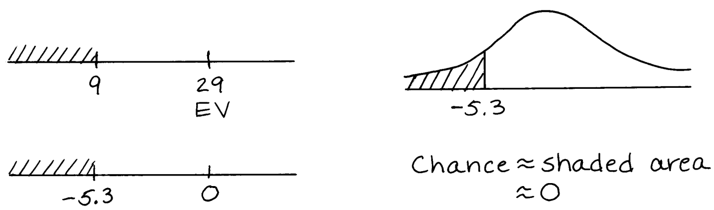
(b) 100 draws are made at random without replacement from a box with 102 1's and 248 0's, where 1 = woman and 0 = man. The expected number of 1's is 29. If the draws are made with replacement, the SE for the number of 1's is $\sqrt{100} \times \sqrt{0.29 \times 0.71} \approx 4.54$. The correction factor is

$$\sqrt{\frac{350 - 100}{350 - 1}} \approx 0.846.$$

The SE for the number, when drawing without replacement, is

$0.846 \times 4.54 \approx 3.8.$

So $z \approx (9 - 29)/3.8 \approx -5.3$. The chance is about 0.



(c) Judge Ford was not choosing at random; he was excluding women.

Comment. After Hans Zeisel made a statistical analysis of Judge Ford's procedures, the percentage of women jurors went up quite dramatically; see note 15 to chapter 26.

7. Disagree. There are $580 + 442 = 1022$ subjects. With a coin, the expected number in the control group is 511, and the SE is about 16, so the chance of getting 442 or fewer in the control group is practically 0. Both patients and doctors know that if you turn up on an odd day of the month, you get the therapy. There may be a considerable temptation to enroll more patients on odd days, and that seems to be what happened.

   Comment. The danger is that the patients enrolled on the odd days will be different from the ones enrolled on the even days. For example, the doctors may tend to enroll relatively healthy patients—who need the therapy less—on the even days. If the control group starts off healthier than the treatment group, the study is biased against the treatment. Tossing a coin is wiser.

8. Model: there is one ticket in the box for each person in the county, age 18 and over. The ticket shows that person's educational level. The data are like 1000 draws from the box.

   Null: the average of the box is 13 years.
   Alt: the average of the box isn't 13 years.

   The expected value for the average of the draws is 13 years, based on the null. The SD of the box is unknown (there is no reason the spread in the county should equal the spread in the nation), but can be estimated as 5 years—the SD of the data. On this basis, the SE for the sample average is estimated as 0.16 years. The observed value for the sample average is 14 years, so

   $z = (\text{obs} - \text{exp})/\text{SE} = (14 - 13)/0.16 \approx 6,$

   and $P \approx 0$. This is probably a rich, suburban county, where the educational level would be higher than average.

9. Something is wrong. The EV for the sum is 20; the SE for the sum is 4. The average of 144 sums should be around 20; the SE for the average is 0.33. The observed value is 3.4 SEs away from expected. That is too many SEs.

You can also get the EV and SE another way: the average of 144 sums is like the sum of $144 \times 100 = 14{,}400$ draws from the original box, divided by 144.

10. Null: the 3 Sunday numbers are like 3 draws made at random (without replacement) from a box containing all 25 numbers in the table. The average of these numbers is nearly 436, and their SD is just about 40. The EV for the average is 436, and the SE is 22; we are using the correction factor here. The 3 Sunday numbers average about 357, so

$$z = (\text{obs} - \text{exp})/\text{SE} = (357 - 436)/22 \approx -3.6,$$

and $P \approx 2/10{,}000$.

Comments. (i) Many deliveries are induced, and some are surgical, so obstetricians really can influence the timing.
(ii) In all, there are 25!/(3!×22!)= 2300 samples of size 3. The 3 Sunday numbers are 344, 377, 351. The number for Saturday, August 6, is also 377; and all the other numbers in the table are larger. Therefore, exactly 2 samples have an average equal to the Sunday average, and no sample average is smaller. The exact significance probability is $2/2300 \approx 9/10{,}000$, compared to 2/10,000 from the curve.

Sundays are for golf.

11. (a) The box has one ticket for each household in Atlanta in June 2003. If the income is over \$52,000, the ticket is marked 1; otherwise, 0. The null hypothesis says that the percentage of 1's in the box is 50%; the alternative, that percentage of 1's in the box is bigger than 50%. The data are like 750 draws from the box.
    (b) The SE for the number of 1's in the sample is $\sqrt{750} \times \sqrt{0.50 \times 0.50} \approx 13.7$; the SE for the percentage of 1's in the sample is 13.7/750, or 1.83%. So $z = 6/1.83 \approx 3.3$, and P is very small.
    (c) Median household income went up.

12. (a) There are 59 pairs, and in 52 of them, the treatment animal has a heavier cortex. On the null hypothesis, the expected number is $59 \times 0.5 = 29.5$ and the SE is $\sqrt{59} \times 0.5 \approx 3.84$. So 52 is nearly 6 SEs above average, and the chance is close to 0. Inference: treatment made the cortex weigh more.
    (b) The average is about 36 milligrams and the SD is about 31 milligrams. The SE for the average is 4 milligrams, so $z = 36/4 = 9$ and $P \approx 0$. (This is like the tax example in section 1.) Inference: treatment made the cortex weigh more.
    (c) This blinds the person doing the dissection to the treatment status of the animal. It is a good idea, because it prevents bias; otherwise, the technician might skew the results to favor the research hypothesis.

Chapter 27. More Tests for Averages

1. No. The expected number of positives is 250, and the SE is $\sqrt{500} \times 0.5 \approx 11$. The observed number is 2.4 SEs above the expected. (Here, a one-sample test is appropriate.)

2. (a) The SE for the difference is 5.9%, so $z = 2.4/5.9 \approx 0.4$; looks like chance.
   (b) The SE for the difference is 1.8; the observed difference is 4.9; so $z = 4.9/1.8 \approx 2.7$ and $P \approx 0.3$ of 1%. The difference looks real.

   Comment. There is more information in the sample average than in the number of positive terms, at least for this example.

3. There are two samples, you need to make a two-sample z-test. Model: there are two boxes. The 2005 box has a ticket for each person in the population, marked 1 for those who would rate clergymen "very high or high," and 0 otherwise. The 2005 data are like 1000 draws from the 1985 box. The 2000 box is set up the same way. The null hypothesis says that the percentage of 1's in the 2005 box is the same as in the 2000 box. The alternative hypothesis says that the percentage of 1's in the 2005 box is smaller than the percentage of 1's in the 2000 box.

   The SD of the 2005 box is estimated from the data as $\sqrt{0.54 \times 0.46} \approx 0.50$. On this basis, the SE for the 2005 number is $\sqrt{1000} \times 0.50 \approx 16$: the number of respondents in the sample who rate clergymen "very high or high" is 540, and the chance error in that number is around 16. Convert the 15 to percent, relative to 1000. The SE for the 2005 percentage is estimated as 1.6%. Similarly, the SE for the 2000 percentage is about 1.5%.

The SE for the difference is computed from the square root law (p. 502) as

$$\sqrt{1.6^2 + 1.5^2} \approx 2.2\%.$$

The observed difference is $54 - 60 = -6\%$. On the null hypothesis, the expected difference is 0%. So z = (obs − exp)/SE $= -6/2.2 \approx -2.7$, and $P \approx 3/1000$. The difference is real. What the cause is, the test cannot say.

Comment. Either a one-sided or a two-sided test can be used. Here, the distinction is not so relevant: for discussion, see chapter 29.

4. You need more information. The method of section 2 does not apply because you do not have two independent samples. The method of sections 3–4 does not apply because you observe two responses for each person. See p. 517.

5. This is like the radiation-surgery example in section 4. Each subject has two possible responses, one to item A and one to item B. The investigators only observe one of the two, chosen at random. To make the test, pretend you have two independent random samples. With item A, the percentage who answer "yes" is 46%; the SE for this percentage is 3.5%. With item B, the percentage is 88% and the SE is 2.4%. The difference between the percentages is $46\% - 88\% = -42\%$. The SE for the difference is conservatively estimated as $\sqrt{3.5^2 + 2.4^2} \approx 4.2\%$. So $z = -42/4.2 = -10$. The framing of the question makes a difference. That is what the experiment tells you.

6. This is just like the previous exercise. In the calculator group, 7.2% get the right answer; in the pencil-and-paper group, 23.6%. The SEs are 1.6% and 2.7%. The difference between the percentages is −16.4%, and the SE for the difference is conservatively estimated as $\sqrt{1.6^2 + 2.7^2} \approx 3.1\%$. So $z = -16.4/3.1 \approx -5$, and $P \approx 0$. The difference is real. (Students who used the calculator seemed to forget what the arithmetic was all about.)

7. (a) This is like the radiation-surgery example in section 4. (Also see review exercises 5 and 6 above.) The SE for the treatment percent is 2.0%. The SE for the control pecent is 4.0%. The SE for the difference is 4.5%. The observed difference, 1.1%, is only 0.24 of an SE. This could easily be due to chance. Income support was a good idea that didn't work.

   (b) This is like example 4. The SE for the treatment average is 0.7 weeks; for the control average, 1.4 weeks; for the difference, 1.6 weeks. The observed difference is −7.5 weeks. So $z = -7.5/1.6 \approx 4.7$ and $P \approx 0$. The difference is real. Income support makes the released prisoners work less, which might explain the findings in part (a).

8. The data can be summarized as follows:

	Prediction Request	Request only
Predicts	22/46	NA
Agrees	14/46	2/46

(a) This is like the radiation-surgery example in section 4. The two percentages are 47.8% and 4.4%. The SEs are about 7.4% and 3.0%, respectively. The difference is 43.4% and the SE for the difference is 8%. So $z = 43.4/8 \approx 5.4$ and $P \approx 0$. The difference is real. People overestimate their willingness to do volunteer work.

(b) The two percentages are 30.4% and 4.4%. The SEs are about 6.8% and 3.0%, respectively. The difference is 26% and the SE for the difference is 7.4%. So $z = 26/7.4 = 3.5$ and $P \approx 2/10{,}000$. The difference is real. Asking people to predict their behavior changes what they will do.

(c) Here, a two-sample z-test is not legitimate. There is only one sample, and two responses for each person in the sample. Both responses are observed, so the method of section 4 does not apply. The responses are correlated, so the method of example 3 does not apply. See p. 517.

Comment. In parts (a) and (b), the number of draws is small relative to the number of tickets in the box. So there is little difference between drawing with or without replacement, and little dependence between the treatment and control averages. See pp. 510, 517.

9. This is like the radiation-surgery example in section 4. In the positive group, the percentage accepted is $28/53 \times 100\% \approx 52.8\%$; in the negative group, 14.8%. The SEs are 6.9% and 4.8%. The difference is 38% and the SE for the difference is 8.4%. So $z = 38/8.4 \approx 4.5$ and $P \approx 0\%$. There is a big difference between the two groups, and the difference cannot be explained by chance. Journals prefer the positive articles.

10. This test is not legitimate. There is dependence between the first-borns and second-borns.

11. (i) The expected value for the difference between the average score in the 2004 sample and the average score in the 1990 sample equals the expected value for the average score in the 2004 sample, minus the expected value for the average score in the 1990 sample. (ii) The expected value for the average score in the 2004 sample is the average of the 2004 box; likewise, the expected value for the average score in the 1990 sample is the average of the 1990 box. (iii) If you put the previous points together, the expected value for the difference between the averages of the two samples equals the difference between the averages of the two boxes. (iv) The null hypothesis says that the difference between the averages of the two boxes is 0. That is why 0 is the right benchmark in the numerator of the z-statistic.

Chapter 28. The Chi-Square Test

1. (a) (i) (b) (iii). See exercises 3–6 on pp. 539–40.

2. Use the method of sections 1–2.

Observed	Expected
1	17.6
10	30.1
16	7.4
35	6.9

$\chi^2 \approx 150$ on 3 degrees of freedom, so $P \approx 0$ and option (ii) is right.

Comments. (i) Judges prefer well-educated grand jurors.
(ii) The expecteds do not have to be whole numbers. For instance, if you roll a die 100 times, the expected number of aces is 16.666 ....

3. Use the method of section 4. The expecteds are as follows (row and column sums are a bit off due to rounding).

	Married	Widowed, divorced, or separated	Never married
Employed	772.0	103.3	221.7
Unemployed	66.2	8.9	19.0
Not in labor force	28.9	3.9	8.3

$\chi^2 \approx 14.2$ on 4 degrees of freedom, which is off the end of the table. By computer, $P \approx 0.7$ of 1%. This is not chance variation. The married men do better at getting jobs. (Or, men with jobs do better at getting married: the χ^2-test will not tell you which is the cause and which is the effect.)

4. (a) Chance: it's a probability histogram.
   (b) The chance that $5 \leq \chi^2 < 5.2$, where χ^2 is computed from 60 rolls of a fair die.
   (c) The chance that $5 \leq \chi^2 < 5.2$ is bigger than the chance that $4.8 \leq \chi^2 < 5$. (The block is bigger.)
   (d) 10%.

   Comment. The exact probability distribution of χ^2, with 60 rolls, is quite irregular. As the number of rolls goes up, the histogram gets closer to the curve. See p. 41 of this manual.

5. (a) With 10 degrees of freedom, P will be bigger. Reason: that curve has more area to the right of 15.
   (b) The P-value is bigger when $\chi^2 = 15$. Reason: the area under the curve to the right of 15 is bigger than the area to the right of 20.

6. Use the method of section 3. The observed frequencies are too close to the expected ones for comfort: $\chi^2 \approx 2$ on 10 degrees of freedom, so $P < 1\%$ (left tail); by computer, $P \approx 0.4$ of 1%. This individual seems to have very good control over the dice. Maybe you should decline his invitation to play craps.

7. Use the method of sections 1–2: $\chi^2 \approx 0.2$ on 2 degrees of freedom, $P \approx 90\%$, a good fit.

8. Make a χ^2-test, as in sections 1–2. We are interested in the chances, not just the average: $\chi^2 \approx 2.6$ on 5 degrees of freedom, $P \approx 75\%$, a good fit.

9. Use the method of section 4. The expecteds are as follows:

	20–24	25–29
Never married	34.9	32.1
Married	25.5	23.5
W/D/S	3.6	3.4

$\chi^2 \approx 17$ on 2 degrees of freedom; by computer, $P < 2/1000$. This is not chance variation. It takes time to get married, especially in Montana.

10. (a) Use the method of section 4 to compute χ^2.

	Observed		Expected	
	Protestant	Catholic	Protestant	Catholic
Acquitted	8	27	6.56	28.44
Convicted	7	38	8.44	36.56

	Obs − Exp	
	Protestant	Catholic
Acquitted	1.44	−1.44
Convicted	−1.44	1.44

Now

$$\chi^2 = \frac{1.44^2}{6.56} + \frac{1.44^2}{8.44} + \frac{1.44^2}{28.44} + \frac{1.44^2}{36.56}$$
$$= 1.44^2 \times \left(\frac{1}{6.56} + \frac{1}{8.44} + \frac{1}{28.44} + \frac{1}{36.56}\right) = 0.69$$

and $P \approx 60\%$. The mistake, apparently, was to compute

$$1.44^2 \div \left(\frac{1}{6.56} + \frac{1}{8.44} + \frac{1}{28.44} + \frac{1}{36.56}\right) = 6.22.$$

(b) Presumably, the defense was thinking that accused persons are convicted independently, with a common probability—except that there is one probability for Catholics and another for Protestants. This model does not seem well related to the criminal justice system, where the facts vary from one case to the next, and some cases involve multiple defendants.

Chapter 29. A Closer Look at Tests of Significance

1. (a) True (p. 547). (b) False (pp. 552–53). (c) False (p. 545).

   Of course, P = 4.7% gives you "statistical significance," and improves the odds of journal publication.

2. Question (i): see p. 562.

3. False. You have to take the sample size into account too. For example, suppose the first investigator gets an average of 52, and the second one gets an average

of 51. The first investigator gets $z = (52 - 50)/1 = 2$ and $P \approx 5\%$. The second investigator gets $z = (51 - 50)/0.33 = 3$ and $P \approx 0.3$ of 1%. (The P-values are two-sided.)

4. Yes: data snooping (p. 547).

5. It is hard to make sense out of "statistical significance" here, because there is no reasonable chance model for the data. The inner planets do not form a sample, they are the inner planets; similarly for the outer ones. (See exercise 2 on pp. 558–59; but see note 20 to chapter 29.)

6. There may be a big effect which is poorly estimated (pp. 552–53). Also, there may be problems in setting up a box model here.

7. The concept of statistical significance does not apply very well, because the data are for the whole population, rather than a sample (p. 556). The difference is practically significant. The center of population is shifting to the West, and that makes a lot of difference to the economy and to the political balance of the country.

8. (a) The question makes sense: the data are from probability samples.
   (b) No. You need to use the half-sample method (section 22.5).
   (c) Yes. Use the method of example 3 on p. 505. The SE for the 2005 sample percentage is 0.22 of 1%, and the SE for 1985 is about the same. The SE for the difference is 0.31 of 1%, so $z = 9/0.31 \approx 29$, and $P \approx 0$. This difference is off the chance scale. Increasing participation by women in the labor force is of great practical importance too.

9. (a) The question makes sense, and the difference in attitudes is important. (This is a practical judgment, not a statistical one.)
   (b) The question makes sense, because the data are based on probability samples; but to answer it, you need to use the half-sample method (section 22.5).
   (c) Now this is like example 3 on p. 505. The SE for the 2000 percentage is 1.4%; for 1970, the SE is 1.5%. The difference is 38%, and the SE for the difference is 2.1%, so $z = 38/2.1 \approx 18$ and $P \approx 0$.

10. $P \approx 5.9\%$ is pretty weak evidence; two-tailed, $P \approx 11.8\%$, which is worse. Even to get these P-values, some data-snooping was needed. The argument is not good.

    Comment. There were also serious problems with the model; see note 38 to chapter 29.

11. The question makes sense, but cannot be answered with the information given: you observe two correlated responses for each subject (p. 517).

12. The conclusion (there are genetic differences among ethnic groups) seems right, but the quotation is overheated.

(i) P-values may be objective, but the 5% line is somewhat arbitrary (pp. 545–46).

(ii) Differences which are statistically significant can be due to chance (p. 547); the P-value only tells you that chance has to work hard to make such a big difference.

(iii) A "statistically significant" difference may be practically insignificant (pp. 552–53).

Chapter 29. Special Review Exercises

1. (a) This is an observational study. It is the prisoners themselves who decide whether to stay in the program or drop out.

   (b) The treatment group consists of the prisoners who finish boot camp. The control group consists of those who drop out.

   (c) Those who stayed the course might have been quite different to start with—better motivated, more self-disciplined—than those who dropped out.

   (d) (i) The treatment group consists of all those who volunteer for the program, whether or not they complete it. The control group consists of those who do not volunteer.
   (ii) The recidivism rate in the two groups is similar, suggesting that the program has little effect.
   Comments. (i) If anything, this comparison is biased in favor of treatment, because those who volunteer would seem more likely to succeed in civilian life than those who do not volunteer. (ii) Completion seems to be an effective screening device: prisoners who finish are motivated and have enough self-discipline to go straight when they are released.

   (e) Most completed. The recidivism rate for those who completed is 29%. The rate for those who did not complete is 74%. The rate for the combined group is 36%. If most dropped out, the rate for the combined group would be nearly 74%. On the other hand, if most completed, the rate for the combined group would be nearly 29%—and it is.

   A more exact answer is possible. Let x be the fraction who completed. Then $29x + 74(1 - x) = 36$, so $x \approx 0.84$.

2. Not a good explanation. The comparison is between left handed ball players and right handed players. A confounder would have to be associated with left handedness—among the players—and cause higher mortality rates. Also see the comment on special review exercise 10, chapter 6.

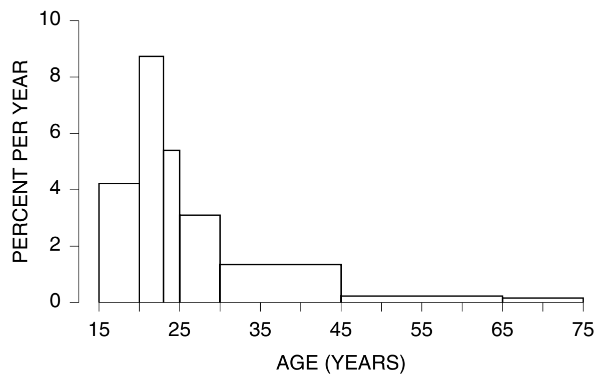
   This study had problems, including a small sample size.

3. False. This is like the graduate admissions study, pp. 17ff. As it turns out, most of the Catholics go to the Secondary Schools, and students in those schools do rather poorly on the proficiency tests.

4. The figure is not a histogram: the class intervals are unequal. If you adjust for the lengths of the intervals, the pattern goes away—as shown by the histogram on the next page. Class intervals include the left endpoint but not the right. For

instance, the rectangle whose base is the interval from 25 to 30 represents the students age 25–29; those age 30 are in the next rectangle. We started the first rectangle at 15, and ended the last one at 75, somewhat arbitrarily. What the histogram shows is that most students are in their early 20s, with a sprinkling of precocious youngsters—and some determined oldsters.

Histogram of ages, special review exercise 4, chapter 29



5. The data are cross-sectional not longitudinal. An alternative explanation: death rates are higher among people who drink, smoke, and don't eat breakfast. So, fewer of these people survive past 65 and get interviewed. The conclusion may be right, but doesn't follow from the data.

6. The difference between the 90th and 50th percentiles is bigger. There is a long right hand tail at work.

7. There is a couple where both husband and wife are about 2 years old. There is a man aged 55 married to a 5-year-old. There is a woman aged 30 married to a 5-year-old. And many other odd couples. Something is wrong.

   Comment. For the motivation, see the discussion of review exercise 11 in chapter 11.

8. $y = 0.533x + 1.667$

9. (a) Investigator A gets the higher correlation: B's correlation is attenuated due to restriction of range. (See exercises 4 and 5 on p. 130, exercise 9 on p. 144, exercise set B on pp. 145–46.)

   Comment. In the March 2005 Current Population Survey, the correlations were about 0.31 and 0.23, respectively.

   (b) The "ecological" correlation—for the state averages—will be higher. See section 9.4.

10. Option (iii) is right—regression effect.

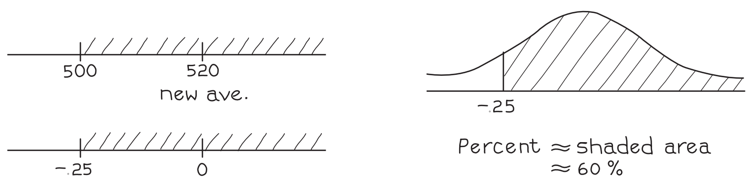
11. A first-year GPA of 3.5 is 1 SD above average. Students with this GPA averaged about $r \times 1 = 0.4$ SDs above average in second year, by the regression method.

Sally must have been above average on second-year GPA by about 0.4 SDs, putting her in the 66th percentile.

12. Something is wrong. The r.m.s. error has to be less than the SD.

13. Less than. The diagram is heteroscedastic, with less scatter around the regression line for women with lower educational levels (p. 192).

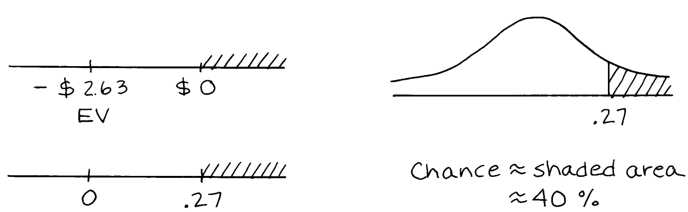
14. The students who scored 500 on the V-SAT averaged about $0.6 \times 500 + 220 = 520$ on the M-SAT. That is the new average. The new SD is the r.m.s. error of the regression line, which is 80 points. Now $(500 - 520)/80 = -0.25$. The area under the normal curve to the right of -0.25 is about 60%. The answer: about 60% of 50,000 = $0.60 \times 50{,}000 = 30{,}000$ students scored better than 500 on the M-SAT.



15. (a) $4/52 \times 3/51 \times 2/50 \approx 2/10{,}000$.
    (b) $48/52 \times 47/51 \times 46/50 \approx 78\%$.
    (c) $36/52 \times 35/51 \times 34/50 \approx 32\%$.
    (d) $100\% - 32\% = 68\%$.

16.
$$\binom{6}{3}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^3 = \frac{20}{46{,}656} \approx \frac{4}{10{,}000}$$

17. The net gain is like the sum of 100 draws from a box with 18 tickets marked "+\$1," 18 tickets marked "−\$1," and 2 tickets marked "−\$0.50." (These last 2 tickets correspond to 0 and 00, where you only lose half your stake.) The average of this box is $-\$1/38 \approx -\$0.0263$. The expected net gain is $-\$2.63$. The SD of the box—don't use the short-cut—is \$0.980. The SE for the net gain is \$9.80. Now $\$2.63/\$9.80 \approx 0.27$. The answer is about 40%.



18. Using a telephone survey tends to exclude the homeless or recently homeless. On this basis, the 3% is a little too low.

Comment. People might not want to admit having been homeless; that also makes the 3% too low. On the other hand, if people were homeless, they might remember that as being more recent than it really was. (Vivid experiences tend to be brought forward in memory.) That would make the 3% too high.

19. (a) Each man in the population has had sex with some number of women, perhaps 0. Adding these numbers up gives the total number of female partners for the men. There is a similarly-defined total number of male partners for the women. The two totals must be equal—if a man has a woman as partner, that woman has the man as partner. Since the number of men in the population is about equal to the number of women, the average number of partners must be about the same.

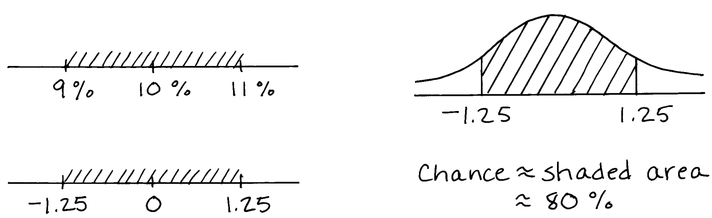
    Comment. Strict equality need not hold—as Laumann points out—although the exceptions seem quite minor. For instance, American men might be more likely than women to have sex while travelling out of the country.

    (b) The total of a list is related to the average not the median, so Laumann seems to have missed the point.

    (c) Lewontin's point is a good one: people in jail, and the homeless, are much more exposed to AIDs than the rest of us. The omission distorts the results in an important way.

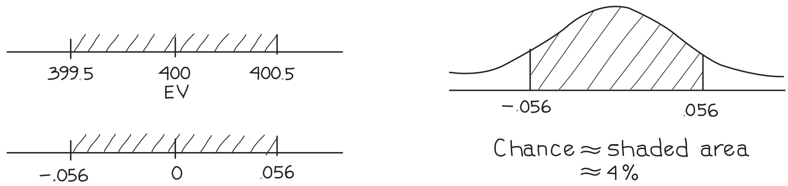
    (d) Non-respondents may well behave quite differently from respondents.

20. The number of sample families without cars is like the sum of 1500 draws from a 0–1 box. There is a ticket in the box for each of the 25,000 families in the town, marked 1 (no car) or 0 (owns cars). So, the fraction of 1's in the box is 0.1, and the SD of the box is $\sqrt{0.1 \times 0.9} = 0.3$. The expected value for the sum is $1500 \times 0.1 = 150$. The SE is $\sqrt{1500} \times 0.3 \approx 12$. The number of sample families without cars will be around 150, give or take 12 or so. Now 12 out of 1500 is 0.8 of 1%. The percentage of sample families without cars will be around 10%, give or take 0.8 of 1% or so. The chance is about 80%.



21. Option (i) is right: section 20.4.

22. The town is large, there is not much difference between drawing with or without replacement. Use the method of section 18.4. The number with phones is like the sum of 500 draws from a box with 80 tickets marked "1" and 20 marked "0." The expected number is $500 \times 0.80 = 400$. The SE is $\sqrt{500} \times \sqrt{0.80 \times 0.20} \approx 8.94$. And $0.5/8.94 \approx 0.056$. From the table, the chance is about 4%.



Comments. (i) By computer, the area under the normal curve between $\pm.056$ is 4.47%.
(ii) You can also use the binomial formula (chapter 15), provided you have a computer to work out the binomial coefficient and the powers:

$$\binom{500}{400}(0.8)^{400}(0.2)^{100} \approx 4.46\%$$

The normal approximation is very good—4.47% vs. 4.46%.

23. (a) Bias; section 19.3.
    (b) A cluster sample is a probability sample, a sample of convenience isn't; sections 19.4, 22.5, 23.4.
    (c) Cluster samples are generally less accurate than simple random samples of the same size, but much cheaper, so they are quite cost effective; section 22.5.

24. False. This is a sample of convenience, not a simple random sample. The formula does not apply (p. 424).

25. (a) (i) (b) (ii) (c) (ii) (iii)

    Comment. The expected value of the *sample average* equals the *population average*. This is so even after the sample is drawn—the expected value is sort of the average over all possible samples, not just the particular sample you happened to draw. In the frequency theory, it is a mistake to say that the expected value of the *population average* equals the *sample average*. See section 21.3 for a similar point about percentages.

26. (a) True: $(488 + 592)/2 = 540$.
    (b) True: just work out the SE and the confidence interval from the SD of 390 and the sample size of 225.
    (c) False. The SD is a large fraction of the average. If the data followed the normal curve, there would be a lot of negative distances travelled.
    (d) True; pp. 411, 418–19.
    (e) False. The population average does not have a probability histogram.
    (f) False. If you double the sample size, you cut the SE for the average not by a full factor of 2, but only by $\sqrt{2} \approx 1.4$. The confidence interval will be about $1/1.4 \approx 0.7$ as wide.

    Comment. The problem is tacitly assuming that the airline used a symmetric confidence interval.

27. (a) Can't be done with the information given, you have to use the half sample method (section 22.5).

    (b) 0.5 of 1%.

28. Disagree. The SE measures the likely size of the chance error. It does not take bias into account. Therefore, neither does the confidence interval.

29. Option (ii) is it. The half sample method must be used (section 22.5).

30. (a) True.

    (b) False. You can use the variability in the data to estimate the SD of the error box, and then compute a standard error for the average. See part (c).

    (c) True. The SD of the data estimates the SD of the error box, hence, the chance error in a single measurement. The SE for the average can be estimated by the square root law, as in section 24.1.

31. (a) 15, the SD of the measurements.

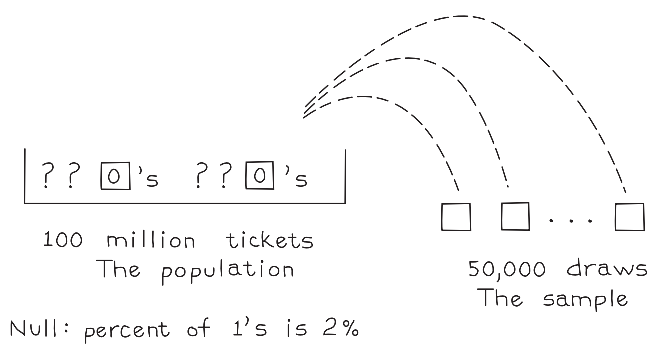
    (b) $15/\sqrt{25}$, the SE for the average.

32. There is a 25% chance for a chick to be *c/c* and have white feathers, so the chance of getting 12 chicks out of 24 with colored feathers is

$$\binom{24}{12} (.75)^{12} (.25)^{12} \approx 0.5 \text{ of } 1\%.$$

You can also use the method of section 18.4, which gives 0.37 of 1%.

33. Model: there is one ticket in the box for each household in the country, marked 1 if the household experienced a burglary within the last 12 months, and 0 otherwise. There are 100 million tickets in this box. The Survey data are like 50,000 draws made at random from the box. (There is essentially no difference between drawing with or without replacement.) If the FBI data are accurate, the percentage of 1's in the box is 2%: this is the null hypothesis. On the null hypothesis, the 3% is higher than the 2% just because of sampling error. The alternative hypothesis: the percentage of 1's in the box is bigger than 2%.



If the null hypothesis is right, the percentage of 1's has an EV of 2%, and the SE is 0.06 of 1%. (The SD of the box should be computed using the 2% specified by the null hypothesis.) The percentage of 1's in the sample is 3%. The difference between 3% and 2% is almost impossible to explain as chance variation: $z \approx 20$. Many burglaries are not reported to the police.

34. (a) True (section 18.4). (b) False (pp. 480–81). (c) False (pp. 480–81).

35. Parts (a–d) can be handled like example 4 on p. 508; parts (e–g), like the radiation-surgery example on pp. 512ff. By convention, the "observed difference" is "treatment − control."
   (a) $z = 0.1/0.13 \approx 0.8$ and $P \approx 21\%$. At baseline, the difference between the two groups is well within the range of chance; the randomization worked. (Compare exercise 3 on p. 511.)
   (b) $z = -3.1/0.15 \approx -21$ and $P \approx 0$. The intervention really got the blood pressure to go down.
   (c) $z = 0.3/0.65 \approx 0.46$ and $P \approx 33\%$. At baseline, the difference between the two groups is well within the range of chance; the randomization worked.
   (d) $z = -4.8/0.69 \approx -7$ and $P \approx 0$. The intervention really got the serum cholesterol to go down.
   (e) $z = 0.3/0.87 \approx 0.34$ and $P \approx 36\%$. At baseline, the difference between the two groups is well within the range of chance; the randomization worked.
   (f) $z = -13.3/0.85 \approx -16$ and $P \approx 0$. The intervention really got them to give up smoking.
   (g) The 6-year death rate in the treatment group was 3.28%, compared to 3.40% in control. The SEs are 0.22 of 1% and 0.23 of 1%. The difference is −0.12 of 1%, and the SE for the difference is 0.32 of 1%. So $z = -0.12/0.32 \approx -0.38$ and $P \approx 34\%$. The intervention got the risk factors down, but didn't change the death rate.

   Comments. (i) The sample sizes used in the calculation are at baseline; no adjustment is made for mortality. (This would be minor.)
   (ii) MRFIT was a well-designed, well-conducted study—which came up with answers contradicting the prevailing wisdom. For more cites to the literature on cholesterol, see note 7 to chapter 29.

36. The question makes sense, but cannot be answered with the information given: you have two correlated responses for each subject (p. 517).

37. (a) Can't be done with the information given, the Current Population Survey isn't a simple random sample (chapter 22).
   (b) Use the method of section 28.4.

Obs		Exp		Obs − Exp	
12	34	22.3	23.7	−10.3	10.3
15	17	15.5	16.5	−0.5	0.5
5	2	3.4	3.6	1.6	−1.6
31	14	21.8	23.2	9.2	−9.2

$\chi^2 \approx 18.2$ on 3 degrees of freedom, $P \approx 4/10{,}000$.

Interpretation: women who are less well educated tend not to be in the labor force; women who are better educated are more likely to be in the labor force, and in professional or managerial jobs.

38. The quote is misleading (pp. 480–81). The test tells you the chance of seeing a big difference, given the null hypothesis. It does not tell you the chance that the null hypothesis is right, given the big difference.

39. These investigators seem to have made a number of statistical errors. For one thing, they did lots of tests; data snooping makes P-values hard to interpret (p. 547): even if all their null hypotheses were right, they were almost bound to find some highly significant differences. For another thing, they seem to be thinking that P measures the size of the effect, and it doesn't (pp. 552–53). The effect they estimated is minute: a 100-fold increase in asbestos fiber concentration only increases the risk of lung cancer by a factor of 1.05. Finally, and most important: smoking is a major cause of lung cancer, and the investigators paid no attention to this variable. The argument is weak, and there is no reason to think that asbestos in the water causes lung cancer.

    Comment. See note 52 to chapter 29.

40. (a) The population and the sample are the same, namely, all Dutch men from two-child families who came of military age in the period 1963–66. The estimates and the parameters coincide, because we have data for the whole population.

    (b) The statistical tests do not make much sense, because the data are for a whole population (p. 556). Even if you think of these men as a sample from larger population, why is it anything like a simple random sample?

    Comments. (i) The first-borns and the second-borns would generally be from different families, due to the design.
    (ii) Belmont and Marolla did interesting work, but the tests seem to be largely ceremonial. See note 20 to chapter 19.