

Chapter 2

1. Sarah performs a CRD with a dichotomous response. She obtains the sampling distribution of the test statistic for Fisher's test for her data; it is given below.

| x | $P(X = x)$ | $P(X \leq x)$ | $P(X \geq x)$ |
|---------|------------|---------------|---------------|
| -0.6667 | 0.0001 | 0.0001 | 1.0000 |
| -0.5278 | 0.0024 | 0.0025 | 0.9999 |
| -0.3889 | 0.0242 | 0.0267 | 0.9975 |
| -0.2500 | 0.1104 | 0.1371 | 0.9733 |
| -0.1111 | 0.2588 | 0.3959 | 0.8629 |
| 0.0278 | 0.3220 | 0.7179 | 0.6041 |
| 0.1667 | 0.2094 | 0.9273 | 0.2821 |
| 0.3056 | 0.0652 | 0.9925 | 0.0727 |
| 0.4444 | 0.0075 | 1.0000 | 0.0075 |

- (a) Find the P-value for the first alternative ($p_1 > p_2$) if $x = 0.1667$.
- (b) Find the P-value for the third alternative ($p_1 \neq p_2$) if $x = -0.2500$.
- (c) Determine **both** the P-value and x that satisfy the following condition: The data are statistically significant but not highly statistically significant for the second alternative ($p_1 < p_2$).
2. Sally performs a CRD with a dichotomous response. She obtains the sampling distribution of the test statistic for Fisher's test for her data; it is given below.

| x | $P(X = x)$ | $P(X \leq x)$ | $P(X \geq x)$ |
|---------|------------|---------------|---------------|
| -0.5000 | 0.0003 | 0.0003 | 1.0000 |
| -0.3920 | 0.0051 | 0.0054 | 0.9997 |
| -0.2841 | 0.0378 | 0.0432 | 0.9946 |
| -0.1761 | 0.1376 | 0.1808 | 0.9568 |
| -0.0682 | 0.2722 | 0.4530 | 0.8192 |
| 0.0398 | 0.3016 | 0.7546 | 0.5470 |
| 0.1477 | 0.1831 | 0.9377 | 0.2454 |
| 0.2557 | 0.0558 | 0.9935 | 0.0623 |
| 0.3636 | 0.0065 | 1.0000 | 0.0065 |

- (a) Find the P-value for the first alternative ($p_1 > p_2$) if $x = 0.1477$.
- (b) Find the P-value for the third alternative ($p_1 \neq p_2$) if $x = -0.1761$.
- (c) Determine **both** the P-value and x that satisfy the following condition: The data are statistically significant but not highly statistically significant for the second alternative ($p_1 < p_2$).
3. Consider a balanced study with six subjects, identified as A, B, C, D, E and G. In the actual study,
- Subjects A, B and C are assigned to the first treatment, and the other subjects are assigned to the second treatment.
 - There are exactly four successes, obtained by A, D, E and G.

This information is needed for parts (a)–(c) below.

- (a) Compute the observed value of the test statistic.
- (b) Assume that the Skeptic is correct. Determine the observed value of the test statistic for the assignment that places C, D and E on the first treatment, and the remaining subjects on the second treatment.
- (c) We have obtained the sampling distribution of the test statistic on the assumption that the Skeptic is correct. It also is possible to obtain a sampling distribution of the test statistic if the Skeptic is wrong *provided* we specify *exactly* how the Skeptic is in error. Assume that the Skeptic is correct about subjects C, D and E, but incorrect about subjects A, B and G.

For the assignment that puts D, E and G on the first treatment, and the other subjects on the second treatment, determine the response for each of the six subjects.

4. Consider an unbalanced study with six subjects, identified as A, B, C, D, E and G. In the actual study,

- Subjects A and B are assigned to the first treatment, and the other subjects are assigned to the second treatment.
- There are exactly two successes, obtained by A and C.

This information is needed for parts (a)–(c) below.

- (a) Compute the observed value of the test statistic.
- (b) Assume that the Skeptic is correct. Determine the observed value of the test statistic for the assignment that places D and E on the first treatment, and the remaining subjects on the second treatment.
- (c) We have obtained the sampling distribution of the test statistic on the assumption that the Skeptic is correct. It also is possible to obtain a sampling distribution of the test statistic if the Skeptic is wrong *provided* we specify *exactly* how the Skeptic is in error. Assume that the Skeptic is correct about subjects A and G, but incorrect about subjects B, C, D and E.

For the assignment that puts D and G on the first treatment, and the other subjects on the second treatment, determine the response for each of the six subjects.

5. Two comparative studies are performed; you are given the following information.

- The observed value of the test statistic is less than 0 in the first study and greater than 0 in the second study.
- The first study is balanced.

I used the website to obtain the exact P-value for Fisher's test for each of the three possible alternatives for each of the two studies, giving me a total of six P-values.

My six P-values are: 0.0061, 0.0123, 0.3560, 0.5572, 0.8346 and 0.9997.

Match each P-value to its alternative and its study.

6. Two comparative studies are performed; you are given the following information.

- The observed value of the test statistic is greater than 0 in the both studies.
- The second study is balanced.

I used the website to obtain the exact P-value for Fisher's test for each of the three possible alternatives for each of the two studies, giving me a total of six P-values.

My six P-values are: 0.0355, 0.0559, 0.3756, 0.7512, 0.8297 and 0.9929.

Match each P-value to its alternative and its study.

7. A comparative study yields the following numbers: $n_1 = 10, n_2 = 20, m_1 = 4$ and $m_2 = 26$. On the assumption the Skeptic is correct, list all possible values of the test statistic.

8. A comparative study yields the following numbers: $n_1 = 20, n_2 = 10, m_1 = 6$ and $m_2 = 24$. On the assumption the Skeptic is correct, list all possible values of the test statistic.

9. **(Extra Credit Problem.)** A balanced CRD with $n = 50$ subjects is performed. The study yields a total of 20 successes. Write an expression for $P(X = 0.00)$ using binomial coefficients; do not compute the answer.

10. **(Extra Credit Problem.)** A balanced CRD with $n = 40$ subjects is performed. The study yields a total of 14 successes. Write an expression for $P(X = 0.30)$ using binomial coefficients; do not compute the answer.

Chapter 3

11. A balanced CRD is performed with a total of 600 subjects. There is a total of 237 successes, with 108 of the successes on the first treatment.

Use the standard normal curve to obtain the approximate P-value for the third alternative, $p_1 \neq p_2$.

12. An unbalanced CRD is performed with a total of 800 subjects. Three hundred subjects are placed on the first treatment and 500 are placed on the second treatment. There is a total of 356 successes, with 126 of the successes on the first treatment. Use the standard normal curve to obtain the approximate P-value for the third alternative, $p_1 \neq p_2$.

13. A sample space has three possible outcomes, B, C, and D. It is known that $P(C) = P(D)$. The operation of the chance mechanism is simulated 10,000 times (runs). The sorted frequencies of the three outcomes (B, C, and D) are:

2322, 2360, and 5318.

- (a) What is your approximation of $P(B)$? To receive credit you must explain your answer.
- (b) What is the best approximation of $P(C)$? To receive credit you must explain your answer.

14. A sample space has four possible outcomes, A, B, C, and D. It is known that $P(A) + P(B) = 0.60$ and $P(C) < P(D)$. The operation of the chance mechanism is simulated 10,000 times (runs). The sorted frequencies of the four outcomes (A, B, C, and D) are:

500, 1528, 2531, and 5441.

Use these simulation results to approximate $P(C)$ and $P(D)$. To receive credit you must explain your answers.

Chapter 5

15. Below are tables (1–3) from three different studies. Match each Table to the correct statement below.

- (a) This table provides evidence that p increased over the course of the study.

- (b) This table provides evidence that p decreased over the course of the study.
- (c) This table provides no evidence that p changed over the course of the study.

Table 1

| Half | <i>S</i> | <i>F</i> | Total |
|------|----------|----------|-------|
| 1st | 30 | 45 | 75 |
| 2nd | 40 | 35 | 75 |

Table 2

| Half | <i>S</i> | <i>F</i> | Total |
|------|----------|----------|-------|
| 1st | 60 | 90 | 150 |
| 2nd | 60 | 90 | 150 |

Table 3

| Half | <i>S</i> | <i>F</i> | Total |
|------|----------|----------|-------|
| 1st | 90 | 110 | 200 |
| 2nd | 50 | 150 | 200 |

Table 4

Current

| Prev. | <i>S</i> | <i>F</i> | Total |
|----------|----------|----------|-------|
| <i>S</i> | 60 | 59 | 119 |
| <i>F</i> | 60 | 80 | 140 |
| Total | 120 | 139 | 259 |

Table 5

Current

| Prev. | <i>S</i> | <i>F</i> | Total |
|----------|----------|----------|-------|
| <i>S</i> | 70 | 50 | 120 |
| <i>F</i> | 49 | 30 | 79 |
| Total | 119 | 80 | 199 |

Table 6

Current

| Prev. | <i>S</i> | <i>F</i> | Total |
|----------|----------|----------|-------|
| <i>S</i> | 18 | 42 | 60 |
| <i>F</i> | 42 | 98 | 140 |
| Total | 60 | 140 | 200 |

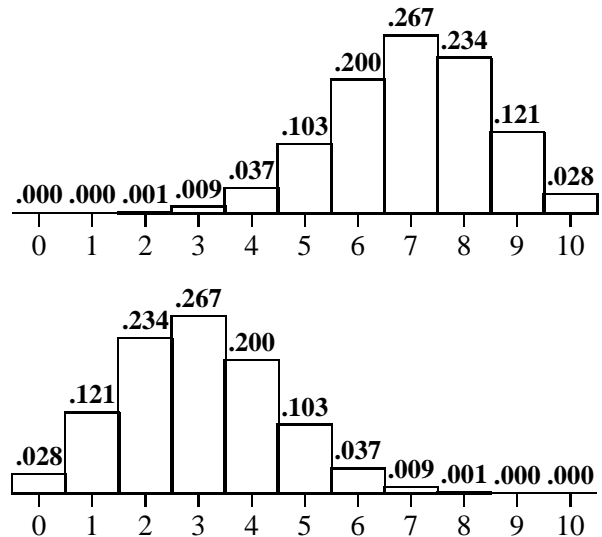
Above are tables (4, 5 and 6) from three different studies. Note that at least one of these tables is the answer to more than of the following four questions.

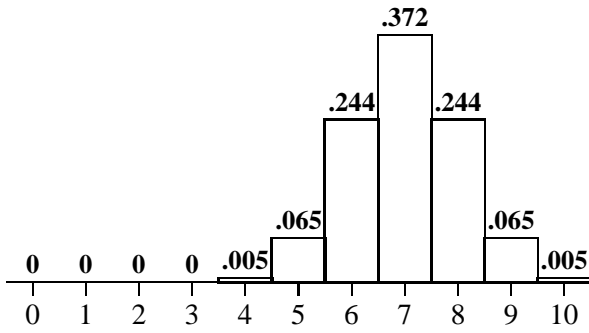
- (d) Which table is for the study that had an F on its first trial and an S on its last trial?
- (e) Which table is for the study in which the subject performed better (S's are good) after an S than after an F?

- (f) Which table is for the study that had an F on its first and last trials?
- (g) Which table is for the study in which the subject performed the same after an S as after an F?
16. On each of four days next week (Monday thru Thursday), Earl will shoot six free throws. Assume that Earl's shots satisfy the assumptions of Bernoulli trials with $p = 0.37$.
- (a) Compute the probability that on any particular day Earl obtains exactly two successes. For future reference, if Earl obtains exactly two successes on any particular day, then we say that the event "Brad" has occurred.
- (b) Refer to part (a). Compute the probability that: next week Brad will occur on Monday and Thursday and will not occur on Tuesday and Wednesday. (Note: You are being asked to compute one probability.)
- (c) Refer to the preamble to this problem. Compute the probability that on any particular day Earl misses his first shot and has a total of two successes.

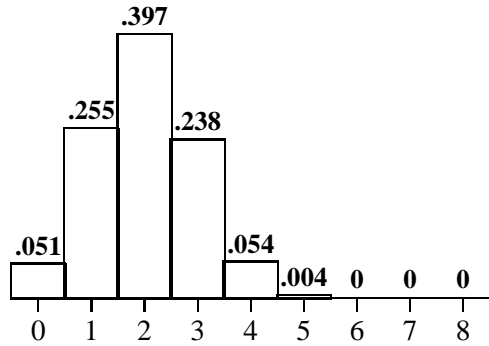
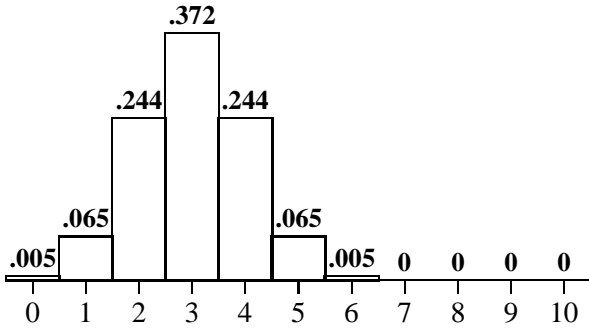
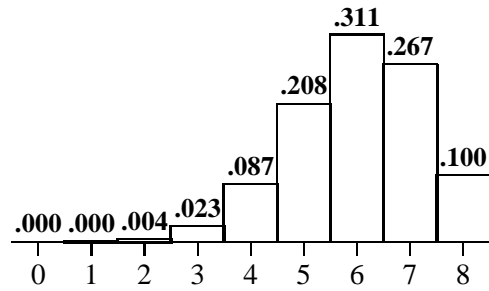
17. On each of four days next week (Monday thru Thursday), Dan will shoot five free throws. Assume that Dan's shots satisfy the assumptions of Bernoulli trials with $p = 0.74$.
- (a) Compute the probability that on any particular day Dan obtains exactly three successes. For future reference, if Dan obtains exactly three successes on any particular day, then we say that the event "Mel" has occurred.
- (b) Refer to part (a). Compute the probability that: next week Mel will occur exactly once and that one occurrence will be on Monday. (Note: You are being asked to compute one probability.)
- (c) Refer to the preamble to this problem. Compute the probability that on any particular day Dan makes his first shot and has a total of four successes.

18. A box contains 14 red cards and six blue cards for a total of 20 cards. Walt is going to select $n = 10$ cards at random *with replacement* from the box. Let W denote the number of red cards that Walt obtains. Let X denote the number of blue cards that Walt obtains. Yale is going to select $n = 10$ cards at random *without replacement* from the box. Let Y denote the number of red cards that Yale obtains. Finally, let Z denote the number of blue cards that Yale obtains. You may use the fact that the probability histograms of the sampling distributions of W , X , Y and Z are pictured below. The number above each rectangle is its height which also equals its area. Note that 0 means zero, whereas .000 means smaller than .0005, but not zero.
- (a) Place a W (X , Y and Z) next to the probability histogram of the sampling distribution of W (X , Y and Z).
- (b) What is the probability that Walt will obtain a representative sample?
- (c) What is the probability that Yale will obtain a sample that is not representative because it has too many red cards?

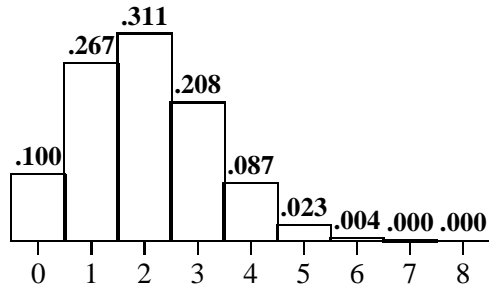




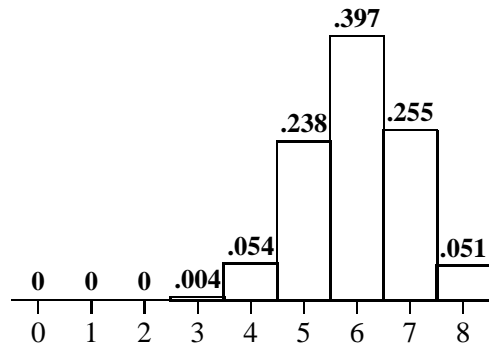
because it has too many red cards?



19. A box contains 15 red cards and five blue cards for a total of 20 cards. Wilma is going to select $n = 8$ cards at random *with replacement* from the box. Let W denote the number of red cards that Wilma obtains. Let X denote the number of blue cards that Wilma obtains. Yolanda is going to select $n = 8$ cards at random *without replacement* from the box. Let Y denote the number of red cards that Yolanda obtains. Finally, let Z denote the number of blue cards that Yolanda obtains. You may use the fact that the probability histograms of the sampling distributions of W , X , Y and Z are pictured to the right. The number above each rectangle is its height which also equals its area. Note that 0 means zero, whereas .000 means smaller than .0005, but not zero.



- Place a W (X , Y and Z) next to the probability histogram of the sampling distribution of W (X , Y and Z).
- What is the probability that Wilma will obtain a representative sample?
- What is the probability that Yolanda will obtain a sample that is not representative



Chapter 6

20. A random sample of size $n = 250$ yields 80 successes. Calculate the 95% confidence interval for p .
21. A random sample of size $n = 452$ yields 113 successes. Calculate the 95% confidence interval for p .
22. George enjoys throwing horse shoes. Last week he tossed 150 shoes and obtained 36 ringers. (Ringers are good.) Next week he plans to throw 250 shoes. Assume that George's tosses satisfy the assumptions of Bernoulli trials.
- Calculate the point prediction of the number of ringers that George will obtain next week.
 - Calculate the 90% prediction interval for the number of ringers George will obtain next week.
 - It turns out that next week George obtains 62 ringers. Given this information, comment on your answers in parts (a) and (b).
23. Bill enjoys throwing darts. Last week he threw 140 darts and obtained 28 bull's-eyes. Next week he plans to throw 350 darts. Assume that Bill's throws satisfy the assumptions of Bernoulli trials.
- Calculate the point prediction of the number of bull's-eyes that Bill will obtain next week.
 - Calculate the 90% prediction interval for the number of bull's-eyes Bill will obtain next week.
 - It turns out that next week Bill obtains 64 bull's-eyes. Given this information, comment on your answers in parts (a) and (b).
24. Carl selects one random sample from a population and calculates three confidence intervals for p . His intervals are below.

| A | B | C |
|---------------------|---------------------|---------------------|
| $\hat{p} \pm 0.080$ | $\hat{p} \pm 0.040$ | $\hat{p} \pm 0.072$ |

Match each confidence interval to its level, with levels chosen from: 80%, 90%, 95%, 98%, and 99%. Note: Clearly, two of these levels will not be used.

25. Carl selects one random sample from a population and calculates three confidence intervals for p . His intervals are below.

| A | B | C |
|----------------------|----------------------|----------------------|
| $\hat{p} \pm 0.0285$ | $\hat{p} \pm 0.0339$ | $\hat{p} \pm 0.0446$ |

Match each confidence interval to its level, with levels chosen from: 80%, 90%, 95%, 98%, and 99%. Note: Clearly, two of these levels will not be used.

Chapter 7

26. An observational study yields the following collapsed table.

| Group | <i>S</i> | <i>F</i> | Total |
|-------|----------|----------|-------|
| 1 | 43 | 57 | 100 |
| 2 | 39 | 61 | 100 |
| Total | 82 | 118 | 200 |

Below are two (partial) component tables for these data. Complete these tables so that Simpson's Paradox is occurring (see Course Notes). Note that there is more than one possible correct answer.

| Subgp A | | | | Subgp B | | | |
|---------|----------|----------|-----|---------|----------|----------|-----|
| Gp | <i>S</i> | <i>F</i> | Tot | Gp | <i>S</i> | <i>F</i> | Tot |
| 1 | 3 | | 20 | 1 | 40 | | 80 |
| 2 | | | 40 | 2 | | | 60 |
| Tot | | | 60 | Tot | | | 140 |

27. An observational study yields the following "collapsed table."

| Group | <i>S</i> | <i>F</i> | Total |
|-------|----------|----------|-------|
| 1 | 72 | 228 | 300 |
| 2 | 88 | 212 | 300 |
| Total | 160 | 440 | 600 |

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring.

| Subgp A | | | | Subgp B | | | |
|---------|----|---|-----|---------|----|---|-----|
| Gp | S | F | Tot | Gp | S | F | Tot |
| 1 | 30 | | 60 | 1 | 42 | | 240 |
| 2 | | | 120 | 2 | | | 180 |
| Tot | | | 180 | Tot | | | 420 |

28. An observational study yields the following collapsed table.

| Group | S | F | Total |
|-------|----|-----|-------|
| 1 | 41 | 59 | 100 |
| 2 | 35 | 65 | 100 |
| Total | 76 | 124 | 200 |

Below are two (partial) component tables for these data. Explain why Simpson's Paradox *cannot* occur for these data.

| Subgp A | | | | Subgp B | | | |
|---------|---|---|-----|---------|----|---|-----|
| Gp | S | F | Tot | Gp | S | F | Tot |
| 1 | 5 | | 20 | 1 | 36 | | 80 |
| 2 | | | 40 | 2 | | | 60 |
| Tot | | | 60 | Tot | | | 140 |

29. An observational study yields the following "collapsed table."

| Group | S | F | Total |
|-------|-----|-----|-------|
| 1 | 48 | 152 | 200 |
| 2 | 52 | 148 | 200 |
| Total | 100 | 300 | 400 |

Below are two component tables for these data. Explain why Simpson's Paradox *cannot* occur for these data. To receive credit, your answer must include computations; i.e. you cannot simply say, "Simpson's Paradox cannot occur because ... cannot happen."

| Subgp A | | | | Subgp B | | | |
|---------|----|---|-----|---------|----|---|-----|
| Gp | S | F | Tot | Gp | S | F | Tot |
| 1 | 12 | | 40 | 1 | 36 | | 160 |
| 2 | | | 80 | 2 | | | 120 |
| Tot | | | 120 | Tot | | | 280 |

Chapter 12

30. A sample of size 40 yields the following sorted data. Note that I have deleted $x_{(39)}$ (the second largest number). This fact will NOT prevent you from answering the questions below.

| | | | | | | |
|------|------|------|------|------|------|------|
| 14.1 | 46.0 | 49.3 | 53.0 | 54.2 | 54.7 | 54.7 |
| 54.7 | 54.8 | 55.4 | 57.6 | 58.2 | 58.3 | 58.7 |
| 58.9 | 60.8 | 60.9 | 61.0 | 61.1 | 63.0 | 64.3 |
| 65.6 | 66.3 | 66.6 | 67.0 | 67.9 | 70.1 | 70.3 |
| 72.1 | 72.4 | 72.9 | 73.5 | 74.2 | 75.3 | 75.4 |
| 75.9 | 76.5 | 77.0 | x | 88.9 | | |

- (a) Construct the density scale histogram for these data using the class intervals: $[10, 40]$, $[40, 50]$, $[50, 60]$, $[60, 75]$ and $[75, 90]$.
- (b) Calculate the first quartile and median of these data.
- (c) Given that the mean of these data is 63.50 and the standard deviation is 12.33, what proportion of the data lie within one standard deviation of the mean?
- (d) Ralph decides to delete the smallest observation, 14.1, from these data. Thus, Ralph has a data set with $n = 39$. Calculate the first quartile and median of Ralph's new data set.
- (e) Refer to (c) and (e). Calculate the mean of Ralph's new data set.
31. A sample of size $n = 16$ yields the following sorted data.

| | | | |
|------|------|------|------|
| 38.7 | 42.1 | 42.7 | 43.7 |
| 45.1 | 46.9 | 49.4 | 50.1 |
| 55.3 | 56.1 | 56.3 | 57.1 |
| 58.0 | 60.6 | y | 62.7 |

Hint: The mean and standard deviation of these data are 51.60 and 7.68.

- (a) How many observations are in the interval $\bar{x} \pm s$?

- (b) Calculate the median and the first quartile of these data.
- (c) Suppose we discover that the observation 38.7 is an error. Recalculate the mean and the first quartile after deleting the observation 38.7.
32. Paula calculates the mean, \bar{x} , and standard deviation, s , of 1,000 measurements. The following are given.
- The interval $\bar{x} \pm 10.0$ contains 450 observations.
 - The interval $\bar{x} \pm 20.0$ contains 680 observations.
 - The interval $\bar{x} \pm 30.0$ contains 830 observations.
 - The standard deviation equals one of the three numbers: 10.0, 20.0, and 30.0.
- (a) Given that the distribution of Paula's data is bell-shaped and symmetric, what is the value of s ?
- (b) Given that the distribution of Paula's data is symmetric with two dominant peaks, what is the value of s ?
- (c) Given that the distribution of Paula's data is strongly skewed, what is the value of s ?
33. Peter calculates the mean, \bar{x} , and standard deviation, s , of 10,000 measurements. The following are given.
- The interval $\bar{x} \pm 40.0$ contains 4500 observations.
 - The interval $\bar{x} \pm 50.0$ contains 6800 observations.
 - The interval $\bar{x} \pm 60.0$ contains 8300 observations.
 - The standard deviation equals one of the three numbers: 40.0, 50.0, and 60.0.
- (a) Given that Peter's s equals 40, describe a likely shape for his data.

- (b) Given that the distribution of Peter's s equals 50, describe a likely shape for his data.
- (c) Given that the distribution of Peter's s equals 60, describe a likely shape for his data.

Chapter 15

34. Independent random samples are selected from two populations. Below are the sorted data from the first population.

| | | | | | |
|-----|-----|-----|-----|-----|-----|
| 362 | 373 | 399 | 428 | 476 | 481 |
| 545 | 564 | 585 | 589 | 590 | 600 |
| 671 | 694 | 723 | 724 | 904 | |

Hint: The mean and standard deviation of these numbers are 571.1 and 144.7.

Below are the sorted data from the second population.

| | | | | | |
|-----|-----|-----|-----|-----|-----|
| 387 | 530 | 544 | 547 | 646 | 766 |
| 786 | 864 | | | | |

Hint: The mean and standard deviation of these numbers are 633.8 and 160.8.

- (a) Calculate Gosset's 90% confidence interval for the mean of the first population.
- (b) Calculate a confidence interval for the median of the second population. Select your confidence level and report it with your answer.
- (c) Suppose that we now learn that the two samples came from the same population. Thus, the two samples can be combined into one random sample from the one population. Use this combined sample to obtain the 95% confidence interval for the median of the population.
35. Independent random samples are selected from two populations. Below are the sorted data from the first population.

53.2 54.2 54.7 55.3 55.9 56.0
 56.3 57.0 58.2 58.5 58.7 61.0
 62.5 62.8 64.4 66.3 67.0 69.0

Hint: The mean and standard deviation of these numbers are 59.50 and 4.80.

Below are the sorted data from the second population.

49.2 53.8 56.9 57.8 58.1
 58.4 62.0 65.4 69.4

Hint: The mean and standard deviation of these numbers are 59.00 and 6.00.

- (a) Calculate Gosset's 90% confidence interval for the mean of the first population.
 - (b) Calculate a confidence interval for the median of the second population. Select your confidence level and report it with your answer.
 - (c) Suppose that we now learn that the two samples came from the same population. Thus, the two samples can be combined into one random sample from the one population. Use this combined sample to obtain the 95% confidence interval for the median of the population.
36. Recall that a confidence interval is *too small* if the number being estimated is *larger* than every number in the confidence interval. Similarly, a confidence interval is *too large* if the number being estimated is *smaller* than every number in the confidence interval.

Each of four researchers selects a random sample from the same population. Each researcher calculates a confidence interval for the mean of the population. The intervals are below.

[24, 41], [30, 39], [20, 33], and [35, 45].

- (a) Nature announces, "Two of the intervals are correct and two are too small." Given this information, what is the narrowest interval that is known to contain the mean? (Hint: The answer is not any of the four confidence intervals.)

- (b) Nature announces, "Two of the intervals are correct, one interval is too small and one interval is too large." Given this information, what is the narrowest interval that is known to contain the mean? (Hint: The answer is not any of the four confidence intervals.)

- (c) It is possible all of these intervals are incorrect. For example, if $\mu = 100$ then every interval is incorrect. But what is the maximum number of these intervals that can be correct? What values of μ will give this maximum number of correct intervals? (Hint: The answer is not any of the four confidence intervals.)

37. Recall that a confidence interval is *too small* if the number being estimated is *larger* than every number in the confidence interval. Similarly, a confidence interval is *too large* if the number being estimated is *smaller* than every number in the confidence interval.

Each of three researchers selects a random sample from the same population. Each researcher calculates a confidence interval for the median of the population. The intervals are below.

[20, 42], [25, 50] and [33, 48].

- (a) Nature announces, "All three intervals are correct." Given this information, determine the narrowest interval that is known to contain the median.
- (b) Nature announces, "Two of the intervals are correct and one interval is too large." Given this information, determine the narrowest interval that is known to contain the median.

Chapter 16

38. Independent random samples are selected from two populations. Below are selected summary statistics.

| Pop. | Mean | Stand. Dev. | Sample size |
|------|-------|-------------|-------------|
| 1 | 62.00 | 10.00 | 17 |
| 2 | 54.00 | 6.00 | 10 |

Calculate s_p .

39. Independent random samples are selected from two populations. Below are selected summary statistics.

| Pop. | Mean | Stand. Dev. | Sample size |
|------|-------|-------------|-------------|
| 1 | 73.00 | 7.00 | 14 |
| 2 | 62.50 | 9.00 | 8 |

Calculate s_p .

40. The null hypothesis is $\mu_X = \mu_Y$. Use Case 1 from Section 16.2 to obtain the P-value for each of the situations described below.

- The alternative is $\mu_X > \mu_Y$; the value of the test statistic is 1.840; the sample sizes are 5 and 5.
- The alternative is $\mu_X < \mu_Y$; the value of the test statistic is -3.150 ; the sample sizes are 6 and 7.
- The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is 1.341; the sample sizes are 5 and 12.

41. The null hypothesis is $\mu_X = \mu_Y$. Use Case 1 from Section 16.2 to obtain the P-value for each of the situations described below.

- The alternative is $\mu_X > \mu_Y$; the value of the test statistic is 1.746; the sample sizes are 6 and 12.
- The alternative is $\mu_X < \mu_Y$; the value of the test statistic is -2.240 ; the sample sizes are 7 and 4.

- The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is 2.910; the sample sizes are 7 and 13.

Chapter 8

42. Below is the table of population counts for a disease and its screening test. (Recall that A means the disease is present and B means the screening test is positive.) On parts (a)–(e) below, report your answers as either: a ratio, like $72/312$ or as a decimal to three digits of precision, for example 0.231.

| | B | B^c | Total |
|-------|-----|-------|-------|
| A | 108 | 12 | 120 |
| A^c | 42 | 698 | 740 |
| Total | 150 | 710 | 860 |

- What proportion of the population is free of the disease?
- What proportion of the population has the disease and would test positive?
- Of those who have the disease, what proportion would test negative?
- What proportion of the population would receive an incorrect screening test result?
- Of those who would receive a correct screening test result, what proportion would receive a correct negative?

43. Below is the table of population counts for a disease and its screening test. (Recall that A means the disease is present and B means the screening test is positive.) On parts (a)–(e) below, report your answers as either: a ratio, like $72/312$ or as a decimal to three digits of precision, for example 0.231.

| | B | B^c | Total |
|-------|-----|-------|-------|
| A | 96 | 12 | 108 |
| A^c | 48 | 564 | 612 |
| Total | 144 | 576 | 720 |

- What proportion of the population has the disease?

- (b) What proportion of the population has the disease and would test negative?
 - (c) Of those who do not have the disease, what proportion would test negative?
 - (d) What proportion of the population would receive a correct screening test result?
 - (e) Of those who would receive an incorrect screening test result, what proportion would receive a false negative?
44. (Hypothetical data.) I have seen 1,000 movies that have been reviewed by Roger Ebert. I enjoyed 72% of these movies and I did not enjoy the remaining 28% of these movies.
- Of the movies I enjoyed, 85% were recommended by Ebert. Of the movies I did not enjoy, 75% were not recommended by Ebert.
- (a) What proportion of these 1,000 movies did Ebert recommend?
 - (b) Suppose that one of these 1,000 movies will be selected at random. Given that Ebert recommended the movie, what is the probability that it is a movie that I enjoyed?
 - (c) I will say that Ebert and I disagreed if either of the following occurs: He recommended it and I did not enjoy it; or I enjoyed it and he did not recommend it. Of these 1,000 movies, for how many did Ebert and I disagree?
- (a) What proportion of the defendants will be convicted by the jury?
 - (b) Given that a defendant is convicted, what is the probability the person is, in fact, guilty?
 - (c) What is the probability that the jury will make a correct decision?
 - (d) Given that the jury makes an incorrect decision, what is the probability that the decision is to release a guilty person?

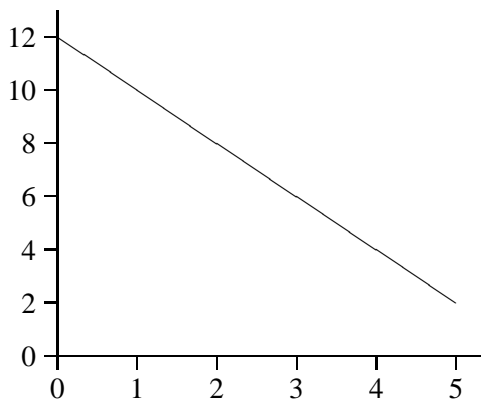
Chapter 13

46. A regression analysis yields the line
- $$\hat{y} = 32 + 0.4x.$$
- One of the subjects, Racheal, has $x = 60$ and $y = 52$. Another subject, Ralph, has $\hat{y} = 56$.
- (a) Calculate Racheal's predicted value, \hat{y} .
 - (b) Calculate Racheal's residual.
 - (c) Calculate Ralph's x .
47. A regression analysis yields the line
- $$\hat{y} = 18 + 0.25x.$$
- One of the subjects, Mary, has $x = 40$ and $y = 32$. Another subject, Mike, has $\hat{y} = 33$.
- (a) Calculate Mary's predicted value, \hat{y} .
 - (b) Calculate Mary's residual.
 - (c) Calculate Mike's x .
48. Fifty students take midterm and final exams. On the midterm exam, the mean score is 45.0 and the standard deviation is 7.00. On the final exam, the mean score is 85.0 with a standard deviation of 14.00. The correlation coefficient of the two scores is 0.64.
- Obtain the least squares regression line for using the final exam score to predict the midterm exam score.

49. Fifty students take two midterm exams. On the first exam, the mean score is 65.0 and the standard deviation is 7.00. On the second exam, the mean score is 55.0 with a standard deviation of 10.00. The correlation coefficient of the two scores is 0.70.

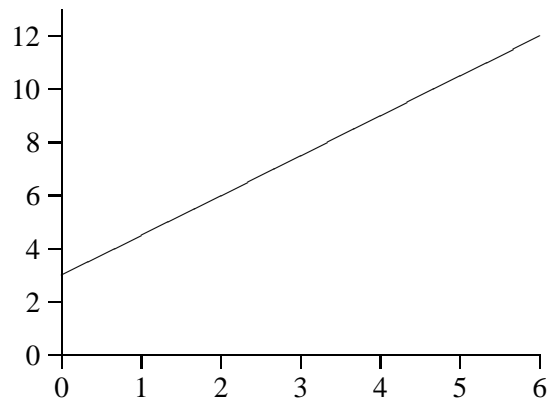
Obtain the least squares regression line for using the second exam score to predict the first exam score.

50. Below is a coordinate system with the regression line $\hat{y} = 12 - 2x$.



- Locate the point that has $x = 3$ and $y = 8$; put an A at that point.
- Locate the point that has $x = 5$ and $e = 2$; put a B at that point.
- Locate the point that has $y = 6$ and $e = -2$; put a C at that point.
- Locate the point that has $\hat{y} = 6$ and $e = -4$; put a D at that point.
- Draw the line that represents all points for which $e = -2$.
- Given that $\bar{x} = 3$, what is the value of \bar{y} ?

51. Below is a coordinate system with the regression line $\hat{y} = 3 + 1.5x$.



- Locate the point that has $x = 1$ and $y = 6$; put an A at that point.
 - Find the value of y for the point that has $x = 2$ and $e = -2$. Put a B at this point.
 - Find the value of x for the point that has $y = 10$ and $e = -2$. Put a C at this point.
 - Draw the line that represents all points for which $e = 4$.
 - Given that $\bar{y} = 7.5$, what is the value of \bar{x} ?
52. Children in grade six take two exams each: one on math and one on vocabulary. For each exam, larger scores are better.

One child, Eric, scores 40 on the math test and 60 on the vocabulary test.

- Eric scored 5 points below the mean on the math exam.
- Eric scored 10 points above the mean on the vocabulary exam.
- Diane obtains the regression line for using the math score to predict the vocabulary score. According to her line, Eric scored 10 points lower than predicted.

Use the above information to obtain Diane's regression line.

53. Children in grade six take two exams each: one on math and one on vocabulary. For each exam, larger scores are better.

One child, Donald, scores 55 on the math test and 75 on the vocabulary test.

- Betty says, “Donald’s two scores are equally good because each score is 5 points above its exam’s mean.
- Debra says, “Donald is better at math; here is why. If you calculate the regression line for using vocabulary score to predict math score, Donald’s actual score on the math exam is 4 points higher than his predicted score.”

Use the above information to obtain the regression line to which Debra refers.

Solutions

1. (a) The P-value is

$$P(X \geq 0.1667) = 0.2821.$$

- (b) The P-value is

$$\begin{aligned} P(X \leq -0.2500) + P(X \geq 0.2500) &= \\ 0.1371 + P(X \geq 0.3056) &= \\ 0.1371 + 0.0727 &= 0.2098. \end{aligned}$$

- (c) The P-value is in the column headed “ $P(X \leq x)$.” The only number in this column that satisfies the stated conditions is 0.0267; thus, 0.0267 is the P-value and it corresponds to $x = -0.3889$.

2. (a) The P-value is

$$P(X \geq 0.1477) = 0.2454.$$

- (b) The P-value is

$$\begin{aligned} P(X \leq -0.1761) + P(X \geq 0.1761) &= \\ 0.1808 + P(X \geq 0.2557) &= \\ 0.1808 + 0.0623 &= 0.2431. \end{aligned}$$

- (c) The P-value is in the column headed “ $P(X \leq x)$.” The only number in this column that satisfies the stated conditions is 0.0432; thus, 0.0432 is the P-value and it corresponds to $x = -0.2841$.

3. (a) The observed value of the test statistic is $x = \hat{p}_1 - \hat{p}_2 = 1/3 - 3/3 = -2/3$.

- (b) The observed value of the test statistic would be $x = \hat{p}_1 - \hat{p}_2 = 2/3 - 2/3 = 0$.

- (c) Remember to go back to the preamble to use what *actually* happened. For the proposed assignment, every subject is moved from where it was in the actual assignment. Thus, the responses of subjects A, B and G will change b/c the Skeptic is incorrect about them, but the responses of C, D and E will not change b/c the Skeptic is correct about them. Thus, B, D, and E will be successes, and A, C and G will be failures.

4. (a) The observed value of the test statistic is $x = \hat{p}_1 - \hat{p}_2 = 1/2 - 1/4 = 0.25$.
- (b) The observed value of the test statistic would be $x = \hat{p}_1 - \hat{p}_2 = 0/2 - 2/4 = -0.50$.
- (c) A will be a S b/c the Skeptic is correct. B will be a S b/c the Skeptic is incorrect. C will be a S b/c it does not change treatment. D will be a S b/c the Skeptic is incorrect. E will be a F b/c it does not change treatment. G will be a F b/c the Skeptic is correct.
5. There are several ways to reach the correct answer. It is generally a good idea to use symmetry, when possible. Because the first study is balanced, its sampling distribution will be symmetric; thus, we begin with the first study.

For the first study, $x < 0$. Define the following quantities.

$$A = P(X \leq x), B = P(X \geq x)$$

$$\text{and } C = P(X \geq -x).$$

By symmetry, $A = C$. The P-values are given below.

| Alt. | P-value |
|--------|--------------|
| $>$ | B |
| $<$ | A |
| \neq | $A + C = 2A$ |

Note that one of the P-values is twice as large as another one. With this in mind, inspect the list of six P-values. The only pair with this property consists of the numbers 0.0061 and 0.0123 (slight rounding error). Thus, $A = 0.0061$ and $2A = 0.0123$.

Next, note (a picture helps here) that $A+B$ must be larger than 1. (Combining A and B we get all the probability to the right of x and all the probability to the left of x and we count the probability at x twice.) Given that $A = 0.0061$, we need $B = 0.9997$.

Next, we consider the second study, which has P-values 0.3560, 0.5572 and 0.8346.

For the second study, $x > 0$. Define the following quantities.

$$D = P(X \leq x), E = P(X \geq x)$$

$$\text{and } F = P(X \leq -x).$$

The P-values are given below.

| Alt. | P-value |
|--------|---------|
| $>$ | E |
| $<$ | D |
| \neq | $E + F$ |

As with the argument earlier, $D + E$ must be larger than 1.

Next, I consider $E + F$. I will prove by contradiction that it equals 0.5572. First, $E + F$ is larger than E ; thus, it cannot be the smallest of the three P-values. If $E + F = 0.8346$, then D and E would be the remaining P-values and they would NOT sum to more than 1.

Given that $E + F = 0.5572$, it follows that E (being smaller) is 0.3560 and D is 0.8346.

6. This problem is nearly identical to the previous one. The actual numbers for the P-values have changed, but that matters little.

We begin with the second study because it is balanced and, hence, has symmetry. The P-value for $>$ is 0.3756 and for \neq is 0.7512, because of the 'doubling' phenomenon. Here the problem gets a bit tricky. The P-value for $<$ could be either 0.8297 or 0.9929. We will need to wait to discover which it is.

For the first study, the P-value for $>$ is the smallest of the three, 0.0355. Thus, the P-value for $<$ must be 0.9929, settling the issue from the second study. This leaves 0.0559 as the P-value for \neq .

7. The possible values of 'a' are: 0, 1, 2, 3 and 4. These yield, respectively:

$$x = 0/10 - 4/20 = -0.20,$$

$$x = 1/10 - 3/20 = -0.05,$$

$$x = 2/10 - 2/20 = 0.10,$$

$$x = 3/10 - 1/20 = 0.25, \text{ and}$$

$$x = 4/10 - 0/20 = 0.40.$$

8. The possible values of 'a' are: 0, 1, 2, 3, 4, 5 and 6. These yield, respectively:

$$x = 0/20 - 6/10 = -0.60,$$

$$x = 1/20 - 5/10 = -0.45,$$

$$x = 2/20 - 4/10 = -0.30,$$

$$x = 3/20 - 3/10 = -0.15,$$

$$x = 4/20 - 2/10 = 0.00,$$

$$x = 5/20 - 1/10 = 0.15, \text{ and}$$

$$x = 6/20 - 0/10 = 0.30.$$

9. The data yield the following table.

| Treat. | S | F | Total |
|--------|----|----|-------|
| 1 | a | b | 20 |
| 2 | c | d | 20 |
| Total | 14 | 26 | 40 |

We want $x = 0.30$ which implies that

$$a/20 - c/20 = 0.30.$$

Multiply both sides by 20 and replace c by $(14 - a)$ to get:

$$a - (14 - a) = 6 \text{ or } 2a = 20,$$

which gives $a = 10$ and $b = 10$. Thus, the desired probability is

$$\frac{\binom{14}{10} \binom{26}{10}}{\binom{40}{20}}$$

10. The data yield the following table.

| Treat. | S | F | Total |
|--------|----|----|-------|
| 1 | a | b | 25 |
| 2 | c | d | 25 |
| Total | 20 | 30 | 50 |

We want $x = 0.00$ which implies that

$$a/25 - c/25 = 0.00,$$

or that $a = c = 10$ and $b = 15$. Thus, the desired probability is

$$\frac{\binom{20}{10} \binom{30}{15}}{\binom{50}{25}}$$

11. The data yield the following table.

| Treatment | S | F | Total |
|-----------|-----|-----|-------|
| 1 | 108 | 192 | 300 |
| 2 | 129 | 171 | 300 |
| Total | 237 | 363 | 600 |

This gives $x = -0.07$, $\sigma = 0.03995$, $z = -1.75$, and $|z| = 1.75$. Thus, the approximate P-value is $2(0.0401) = 0.0802$.

12. The data yield the following table.

| Treatment | S | F | Total |
|-----------|-----|-----|-------|
| 1 | 126 | 174 | 300 |
| 2 | 230 | 270 | 500 |
| Total | 356 | 444 | 800 |

This gives $x = -0.04$, $\sigma = 0.03632$, $z = -1.10$ and $|z| = 1.10$. Thus, the approximate P-value is $2(0.1357) = 0.2714$.

13. (a) B/c C and D have the same probability, their frequencies should be reasonably close. Thus, their frequencies are 2322 and 2360, leaving 5318 as the frequency for B. The approximation to $P(B)$ is 0.5318.
 (b) From (a), we have decided that the frequencies for C and D are 2322 and 2360, but there is no way to know which goes

with C and which with D. I gave full-credit to answers that said to use either of these or their mean. Thus, 0.2322, 0.2360 and 0.2341 all received full-credit.

14. The frequencies for A and B should total a number that is close to 6000. Thus, 500 and 5441 are their frequencies. This leaves 0.1528 and 0.2531 as the approximations to $P(C)$ and $P(D)$, respectively. Note: Many persons lost one-half point for neglecting to put a decimal point in their answers.

15. (a) This describes Table 1.
 (b) This describes Table 3.
 (c) This describes Table 2.
 (d) There will be one subtracted from the column total for F and one subtracted from the row total for S . Thus, the row total for F will be one larger than the column total for F ; only Table 4 satisfies this condition.
 (e) We need $\hat{p}_1 > \hat{p}_2$; only Table 4 satisfies this condition.
 (f) There will be one subtracted from the column total for F and one subtracted from the row total for F . Thus, the row and column totals for F will be the same number; only Table 6 satisfies this condition.
 (g) We need $\hat{p}_1 = \hat{p}_2$; only Table 6 satisfies this condition.

16. (a) This is a binomial problem b/c it is about the total number of successes. The probability of exactly two successes is

$$\frac{6!}{2!4!}(0.37)^2(0.63)^4 = 15(0.1369)(0.1575) = 0.3234.$$

- (b) For this part each day is a trial and the probability that a day yields a success is, from part (a), $p = 0.3234$. We use the multiplication rule b/c the question is about a particular sequence. $P(SFFS) =$

$$0.3234(0.6766)(0.6766)(0.3234) = 0.0479.$$

- (c) This is trickier. Let Y denote the number of successes Earl gets on his last five shots. We want the probability of an F on the first shot AND ($Y = 2$). These probabilities are $q = 0.63$ and

$$\frac{5!}{2!3!}(0.37)^2(0.63)^3 = 0.3423.$$

Finally, by the multiplication rule, we multiply these answers to get

$$0.63(0.3423) = 0.2156.$$

17. (a) This is a binomial problem b/c it is about the total number of successes. The probability of exactly three successes is

$$\frac{5!}{3!2!}(0.74)^3(0.26)^2 =$$

$$10(0.4052)(0.0676) = 0.2739.$$

- (b) For this part each day is a trial and the probability that a day yields a success is, from part (a), $p = 0.2739$. We use the multiplication rule b/c the question is about a particular sequence. $P(SFFF) =$

$$0.2739(0.7261)(0.7261)(0.7261) = 0.1049.$$

- (c) Let Y denote the number of successes Dan gets on his last four shots. We want the probability of a S on the first shot AND ($Y = 3$). These probabilities are $p = 0.74$ and

$$\frac{4!}{3!1!}(0.74)^3(0.26) = 0.4214.$$

Finally, by the multiplication rule, we multiply these answers to get

$$0.74(0.4214) = 0.3118.$$

18. (a) The first picture is for W , the second is for X , the third is for Y and the fourth is for Z .

- (b) Walt will obtain a representative sample if, and only if, he gets three blue cards. From the 2nd picture, $P(X = 3) = 0.267$.
- (c) The event is $Y > 7$. From the 3rd picture,
- $$P(Y > 7) = 0.244 + 0.065 + 0.005 = 0.314.$$
19. (a) The first picture is for W , the third is for X , the fourth is for Y and the second is for Z .
- (b) Wilma will obtain a representative sample if, and only if, she gets two blue cards. From the 3rd picture, $P(X = 2) = 0.311$.
- (c) The event is $Y > 6$. From the 4th picture,
- $$P(Y > 6) = 0.255 + 0.051 = 0.306.$$
20. First, $\hat{p} = 80/250 = 0.320$. The 95% confidence interval is
- $$0.320 \pm 1.96 \sqrt{\frac{0.32(0.68)}{250}} = 0.320 \pm 0.058 = [0.262, 0.378].$$
21. First, $\hat{p} = 113/452 = 0.250$. The 95% confidence interval is
- $$0.250 \pm 1.96 \sqrt{\frac{0.25(0.75)}{452}} = 0.250 \pm 0.040 = [0.210, 0.290].$$
22. (a) First, $\hat{p} = 36/150 = 0.24$. Thus, the point prediction is
- $$m\hat{p} = 250(0.24) = 60.$$
- (b) The 90% prediction interval is $60 \pm$
- $$1.645 \sqrt{60(0.76)} \sqrt{1 + (250/150)} = 60 \pm 18.14 = [42, 78],$$
- after rounding.
- (c) With 62 ringers, the point prediction is too small by 2, but the prediction interval is correct b/c 62 is between 42 and 78.
23. (a) First, $\hat{p} = 28/140 = 0.20$. Thus, the point prediction is
- $$m\hat{p} = 350(0.2) = 70.$$
- (b) The 90% prediction interval is $70 \pm$
- $$1.645 \sqrt{70(0.8)} \sqrt{1 + (350/140)} = 70 \pm 23.03 = [47, 93],$$
- after rounding.
- (c) With 64 bull's-eyes, the point prediction is too large by 6, but the prediction interval is correct b/c 64 is between 47 and 93.
24. Based on the half-widths of the CIs, B has the smallest confidence level and A has the largest. But here is the key. The CIs differ only in which z they use, namely three of the following: 1.282, 1.645, 1.96, 2.326 and 2.576. Note that A is twice as wide as B; by inspection, this implies that A uses 2.576 and B uses 1.282 b/c these are the two z 's such that one is twice as large as the other. Thus, A is 99% and B is 80%. Finally, $0.072/0.080 = 0.9$. Thus, the z for C is 90% of 2.576, i.e. $(0.9)(2.576) = 2.3184$, or, allowing for rounding error, $z = 2.326$ and C is 98%.
25. Based on the half-widths of the CIs, A has the smallest confidence level and C has the largest. But here is the key. The CIs differ only in which z they use, namely three of the following: 1.282, 1.645, 1.96, 2.326 and 2.576. The ratio for B to A is
- $$0.0339/0.0285 = 1.189.$$
- Thus, we need to find two z 's in this ratio. By trial-and-error, the two z 's could be 1.96 and 1.645, or they could be 1.96 and 2.326. The ratio for C to A is
- $$0.0446/0.0285 = 1.565.$$

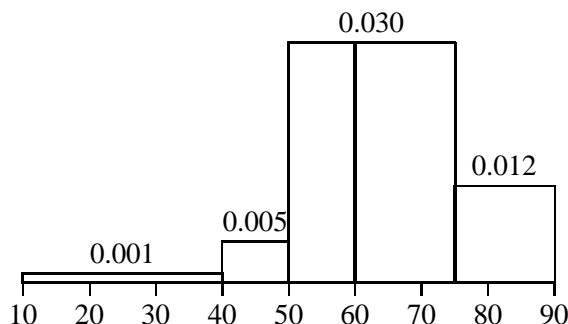
By trial-and-error, the two z 's are 2.576 and 1.645. Thus, the confidence levels are 90%, 95% and 99%.

26. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need $c \geq 7$ in Subgroup A and $c \geq 31$ in Subgroup B. Also, the two c 's must sum to the 39 in the collapsed table. There are two possible answers: 7 and 32; 8 and 31.
27. In the collapsed table, $\hat{p}_1 < \hat{p}_2$. To get a reversal, we need $c \leq 59$ in Subgroup A and $c \leq 31$ in Subgroup B. Also, the two c 's must sum to the 88 in the collapsed table. There are three possible answers: 59 and 29; 58 and 30; 57 and 31.
28. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need $c \geq 11$ in Subgroup A and $c \geq 28$ in Subgroup B. Also, the two c 's must sum to the 35 in the collapsed table. This combination of restrictions is impossible.
29. In the collapsed table, $\hat{p}_1 < \hat{p}_2$. To get a reversal, we need $c \leq 23$ in Subgroup A and $c \leq 26$ in Subgroup B. Also, the two c 's must sum to the 52 in the collapsed table. This combination of restrictions is impossible.

30. (a) First, we create the following table.

| Class Interval | Freq. | Rel. Fr. | Density (rounded) |
|----------------|-------|----------|-------------------|
| 10–40 | 1 | 0.025 | 0.001 |
| 40–50 | 2 | 0.050 | 0.005 |
| 50–60 | 12 | 0.300 | 0.030 |
| 60–75 | 18 | 0.450 | 0.030 |
| 75–90 | 7 | 0.175 | 0.012 |

The density scale histogram is below.



- (b) The median is the mean of the numbers in positions 20 and 21; i.e. 63.0 and 64.3—the median is 63.65. The lower half of the data has 20 observations. The first quartile is the mean of 55.4 and 57.6; i.e. 56.5.

(c) The one sd interval ranges from

$$63.50 - 12.33 = 51.17, \text{ to}$$

$$63.50 + 12.33 = 75.83.$$

This interval contains 32 observations; thus, the proportion within it is $32/40 = 0.80$.

- (d) The median is the number in position 20; i.e. 64.3. The lower half of the data has 19 observations. The first quartile is 57.6.
- (e) The total of the 40 numbers is 40 times the mean, or $40(63.50) = 2540$. After deleting 14.1, the total of the 39 remaining numbers is $2540 - 14.1 = 2525.9$. The mean of the 39 remaining numbers is $2525.9/39 = 64.77$.

31. (a) The one sd interval ranges from

$$51.60 - 7.68 = 43.92, \text{ to}$$

$$51.60 + 7.68 = 59.28.$$

This interval contains 9 observations.

- (b) The median is $(50.1 + 55.3)/2 = 52.7$. The first quartile is $(43.7 + 45.1)/2 = 44.4$.
- (c) After deleting 38.7, the sample size is 15. The first quartile is 45.1.

The total of the original 16 numbers is $16(51.60) = 825.6$. The total of the 15 numbers is $825.6 - 38.7 = 786.9$. The new mean is $786.9/15 = 52.46$.

32. (a) Given the shape of her distribution, the empirical rule should work well. Thus, the one sd interval should contain about 68% of the data, or 680 observations. Thus, $s = 20.0$.

- (b) Given the shape of her distribution, the empirical rule should not work well. In particular, the one sd interval should contain much less than 68% of the data, or many fewer than 680 observations. Thus, $s = 10.0$.
- (c) Given the shape of her distribution, the empirical rule should not work well. In particular, the one sd interval should contain much more than 68% of the data, or many more than 680 observations. Thus, $s = 30.0$.
33. (a) Given his s , the one sd interval contains much less than 68% of the data. Thus, his distribution might be symmetric with two dominant peaks.
- (b) Given his s , the one sd interval contains exactly 68% of the data, just as predicted by the empirical rule. Thus, his shape might be symmetric and bell-shaped.
- (c) Given his s , the one sd interval contains much more than 68% of the data. Thus, his distribution might be strongly skewed.
34. (a) First, $n = 17$; thus, there are 16 degrees of freedom making $t = 1.746$. The CI is
- $$571.1 \pm 1.746(144.7)/\sqrt{17} =$$
- $$571.1 \pm 61.3 = [509.8, 632.4].$$
- (b) B/c $n = 8$ is smaller than 21, we can use the exact CI. From Table A.7 the possible answers are:
- 99.2%: [387, 864], and
 - 93.0%: [530, 786].
- (c) B/c $n = 25$ is larger than 20, we must use the approximate CI. First,
- $$k' = \frac{25 + 1}{2} - \frac{1.96\sqrt{25}}{2} =$$
- $$13 - 4.9 = 8.1.$$
- Thus, $k = 8$ and the CI is [530, 671].
35. (a) First, $n = 18$; thus, there are 17 degrees of freedom making $t = 1.740$. The CI is
- $$59.50 \pm 1.740(4.80)/\sqrt{18} =$$
- $$59.50 \pm 1.97 = [57.53, 61.47].$$
- (b) B/c $n = 9$ is smaller than 21, we can use the exact CI. From Table A.7 the possible answers are:
- 99.6%: [49.2, 69.4],
 - 96.1%: [53.8, 65.4], and
 - 82.0%: [56.9, 62.0].
- (c) B/c $n = 27$ is larger than 20, we must use the approximate CI. First,
- $$k' = \frac{27 + 1}{2} - \frac{1.96\sqrt{27}}{2} =$$
- $$14 - 5.1 = 8.9.$$
- Thus, $k = 8$ and the CI is [56.0, 62.5].
36. (a) $39 < \mu \leq 41$. Note: Throughout this problem you will receive full credit even if you confuse ' $<$ ' with ' \leq .' You will, however, lose credit for writing $\mu = 40$ or 41; i.e. if you assume that μ must be an integer.
- (b) $33 < \mu < 35$.
- (c) The maximum number of correct CIs is 3. It will occur if $30 \leq \mu \leq 33$ or $35 \leq \mu \leq 39$.
37. (a) $33 \leq \nu \leq 42$.
- (b) $25 \leq \nu < 33$.
38. First,
- $$s_p^2 = \frac{16(100) + 9(36)}{16 + 9} = 76.96.$$
- Thus, $s_p = 8.773$.
39. First,
- $$s_p^2 = \frac{13(49) + 7(81)}{13 + 7} = 60.2.$$
- Thus, $s_p = 7.759$.

40. (a) $0.05 < \text{P-value} < 0.10$.
 (b) $\text{P-value} < 0.005$.
 (c) $\text{P-value} = 0.20$.
41. (a) $\text{P-value} = 0.05$.
 (b) $0.025 < \text{P-value} < 0.05$.
 (c) $\text{P-value} < 0.01$.
42. (a) $740/860 = 0.860$.
 (b) $108/860 = 0.126$.
 (c) $12/120 = 0.100$.
 (d) $(12 + 42)/860 = 54/860 = 0.063$.
 (e) $698/(698 + 108) = 698/806 = 0.866$.
43. (a) $108/720 = 0.150$.
 (b) $12/720 = 0.017$.
 (c) $564/612 = 0.922$.
 (d) $(96 + 564)/720 = 660/720 = 0.917$.
 (e) $12/(12 + 48) = 12/60 = 0.200$.
44. A movie is an A if I enjoyed it and a B if Ebert recommended it. I will create the table of population counts; alternatively, you could use population proportions. We are given, or can deduce the following information.
- I enjoyed 720 of the movies.
 - I did not enjoy 280 of the movies.
 - Ebert recommended 85% of the 720 movies I enjoyed; this equals $0.85(720) = 612$ movies.
 - Ebert did not recommend 75% of the 280 movies I did not enjoy; this equals $0.75(280) = 210$ movies.

This information yields the following table.

| | B | B^c | Total |
|-------|-----|-------|-------|
| A | 612 | 108 | 720 |
| A^c | 70 | 210 | 280 |
| Total | 682 | 318 | 1000 |

This table allows us to answer all of the questions.

- (a) $682/1000 = 0.682$.
 (b) $612/682 = 0.897$.
 (c) $108 + 70 = 178$.

45. Let A denote that the defendant is guilty and let B denote conviction. Because I prefer to work with counts rather than proportions, I will set the number of trials at 1000 and proceed as in the previous solution. Alternatively, you could use population proportions. We are given, or can deduce the following information.

- The number of guilty defendants is 820.
- The number of not guilty defendants is 180.
- The jury convicted 75% of the 820 guilty defendants; this equals $0.75(820) = 615$ convicted guilty defendants.
- The jury convicted 40% of the 180 not guilty defendants; this equals $0.40(180) = 72$ convicted not guilty defendants.

This information yields the following table.

| | B | B^c | Total |
|-------|-----|-------|-------|
| A | 615 | 205 | 820 |
| A^c | 72 | 108 | 180 |
| Total | 687 | 313 | 1000 |

This table allows us to answer all of the questions.

- (a) $687/1000 = 0.687$.
 (b) $615/687 = 0.895$.
 (c) $(615 + 108)/1000 = 723/1000 = 0.723$.
 (d) $205/(205 + 72) = 205/277 = 0.740$.

46. (a) Her predicted value is

$$\hat{y} = 32 + 0.4(60) = 56.$$

- (b) Her residual is

$$e = 52 - 56 = -4.$$

(c) His x satisfies

$$56 = 32 + 0.4x,$$

which yields $x = 60$.

47. (a) Her predicted value is

$$\hat{y} = 18 + 0.25(40) = 28.$$

(b) Her residual is

$$e = 32 - 28 = 4.$$

(c) His x satisfies

$$33 = 18 + 0.25x,$$

which yields $x = 60$.

48. The final is X and the midterm is Y . Thus,

$$\bar{x} = 85.0, s_X = 14.00, \bar{y} = 45.0,$$

$$s_Y = 7.00, \text{ and } r = 0.64.$$

Thus,

$$b_1 = 0.64(7/14) = 0.32, \text{ and}$$

$$b_0 = 45 - 0.32(85) = 17.8.$$

The regression line is

$$\hat{y} = 17.8 + 0.32x.$$

49. The second exam is X and the first exam is Y .

Thus,

$$\bar{x} = 55.0, s_X = 10.00, \bar{y} = 65.0,$$

$$s_Y = 7.00, \text{ and } r = 0.70.$$

Thus,

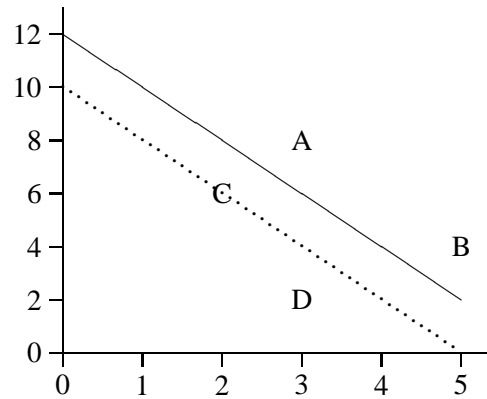
$$b_1 = 0.70(7/10) = 0.49, \text{ and}$$

$$b_0 = 65 - 0.49(55) = 38.05.$$

The regression line is

$$\hat{y} = 38.05 + 0.49x.$$

50. You can identify the points visually or by using algebra. The points are plotted below, followed by the algebraic derivations.



(a) We are given x and y , so we can just plot the point.

(b) Given x , we get $\hat{y} = 2$; given \hat{y} and e , we get $y = 4$.

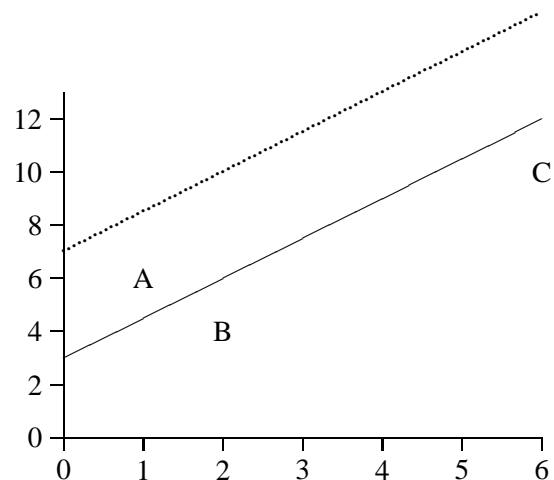
(c) Given y and e , we get $\hat{y} = 8$. Given \hat{y} , we get $x = 2$.

(d) Given \hat{y} and e , we get $y = 2$. Given \hat{y} we get $x = 3$.

(e) The line is parallel to the regression line, a (vertical) distance of two units below it.

(f) Recall that if $x = \bar{x}$, then $\hat{y} = \bar{y}$. Thus, $\bar{y} = 12 - 2(3) = 6$.

51. You can identify the points visually or by using algebra. The points are plotted below, followed by the algebraic derivations.



- (a) We are given x and y , so we can just plot the point.
- (b) Given x , we get $\hat{y} = 6$; given \hat{y} and e , we get $y = 4$.
- (c) Given y and e , we get $\hat{y} = 12$. Given \hat{y} , we get $x = 6$.
- (d) The line is parallel to the regression line, a (vertical) distance of four units above it.
- (e) Recall that if $x = \bar{x}$, then $\hat{y} = \bar{y}$. Thus, $7.5 = 3 + 1.5x$, which yields $x = 3$.

52. Math is X and vocabulary is Y . For Eric, $x = 40$ and $y = 60$. Also, $\bar{x} = 45$ and $\bar{y} = 50$. Finally, Eric's residual for Diane's line is $e = -10$. The regression equation is

$$\hat{y} = b_0 + b_1x.$$

We need to solve this for the two unknowns, b_0 and b_1 . Using the law of the preservation of mediocrity,

$$50 = b_0 + 45b_1.$$

Using information on Eric,

$$70 = b_0 + 40b_1.$$

Subtracting these, we get

$$5b_1 = -20, \text{ or } b_1 = -4.$$

Thus,

$$50 = b_0 - 180, \text{ or } b_0 = 230.$$

Thus, the regression line is

$$\hat{y} = 230 - 4x.$$

53. Math is Y and vocabulary is X . For Donald, $x = 75$ and $y = 55$. Also, $\bar{x} = 70$ and $\bar{y} = 50$. Finally, Donald's residual for Debra's line is $e = 4$. The regression equation is

$$\hat{y} = b_0 + b_1x.$$

Using the law of the preservation of mediocrity,

$$50 = b_0 + 70b_1.$$

Using information on Donald,

$$51 = b_0 + 75b_1.$$

Subtracting these, we get

$$5b_1 = 1, \text{ or } b_1 = 0.2.$$

Thus,

$$50 = b_0 + 14, \text{ or } b_0 = 36.$$

Thus, the regression line is

$$\hat{y} = 36 + 0.2x.$$