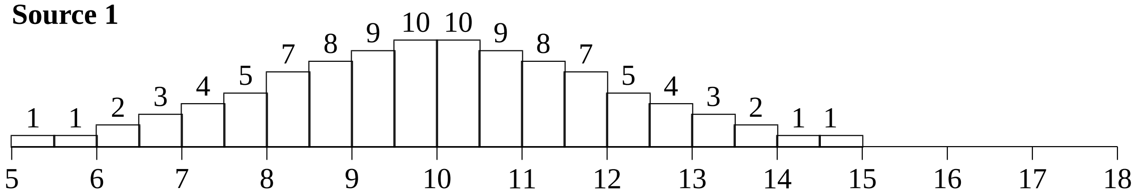
Practice Exam Questions; Statistics 301; Professor Wardrop

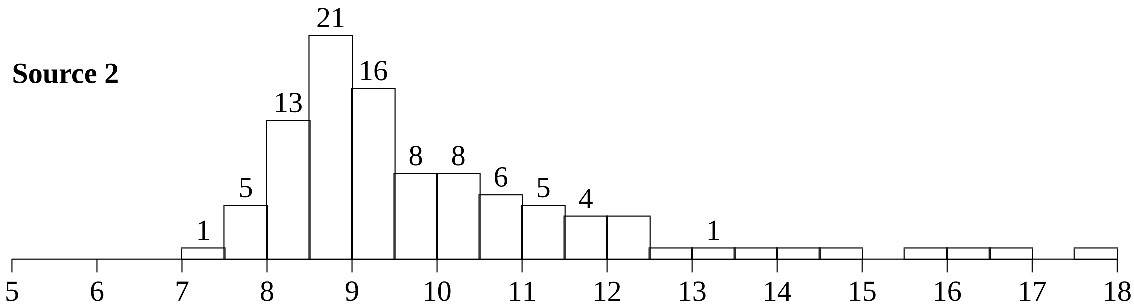
Chapters 1, 12, 2, and 3

1. Measurements are collected from 100 subjects from each of two sources. The data yield the following frequency histograms. The number above each rectangle is its height.

Source 1



Source 2



Each sample has the same mean, 10.00. In order to answer (b) and (c) below, refer to the empirical rule for interpreting s, taking into account the shape of the histogram. Do not try to calculate s because you do not have enough information to do so. In addition, you will receive no credit for simply identifying the correct s; you must provide an explanation.

(a) What is the most precise correct statement that you can make about the numerical value of the median of the data from source 2? **Do not explain your answer.** Hint: Here is a correct statement: The median is between 0 and 20. This statement is not precise enough to receive any credit.

(b) Among the possibilities 1.50, 2.00 and 2.50, which is the numerical value of s for the data from source 1? **Explain your answer.**

(c) Among the possibilities 1.00, 1.50 and 2.00, which is the numerical value of s for the data from source 2? **Explain your answer.**

2. The mean and median of Al's $n = 3$ observations both equal 10. The mean and median of Bev's $n = 5$ observations both equal 18.

(a) Carol combines Al's and Bev's data into one collection of $n = 8$ observations. Can the mean of Carol's data be calculated from the information given? If you think not, just say that. If you think it can, then calculate Carol's mean.

(b) Refer to part (a). Demonstrate, by an explicit example, that there is not enough information to determine Carol's median. Hint: Find two sets of data sets that satisfy Al's and Bev's conditions, yet, when combined, give different medians.

3. A sample of size 40 yields the following sorted data. Note that I have x-ed out $x_{(39)}$ (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) Calculate range, IQR, and median of these data.

(b) Given that the mean of these data is 63.50 (exactly) and the standard deviation is 12.33, what proportion of the data lie within one standard deviation of the mean?

(c) How does your answer to (b) compare to the empirical rule approximation?

(d) Ralph decides to delete the smallest observation, 14.1, from these data. Thus, Ralph has a data set with $n = 39$. Calculate the range, IQR, and median of Ralph's new data set.

(e) Refer to (d). Calculate the mean of Ralph's new data set.

4. Sarah performs a CRD with a dichotomous response and obtains the following data.

Treatment	S	F	Total
1	a	b	18
2	c	d	12
Total	8	22	30

Next, 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 $a = 6$.

(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 ($p_1 < p_2$).

5. Sarah performs a CRD with a dichotomous response and obtains the following data.

Treatment	S	F	Total
1	a	b	22
2	c	d	16
Total	8	30	38

Next, 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 $a = 6$.

(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 ($p_1 < p_2$).

6. 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. These new sampling distributions are used in the study of **statistical power** which is briefly described in Chapter 7 of the text. 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.

7. 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. These new sampling distributions are used in the study of **statistical power** which is briefly described in Chapter 7 of the text. 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.

8. A comparative study is performed; you are given the following information.

   - The total number of subjects equals 33.
   - The observed value of the test statistic is greater than 0.

   I used the website to obtain the exact P-value for Fisher's test for each of the three possible alternatives. These three P-values are below along with three bogus P-values.

Set 1:	0.2450	0.4688	0.9233
Set 2:	0.1445	0.2890	0.9625

(a) Which set contains the correct P-values: 1 or 2? (No explanation is needed.)

(b) For the set you selected in part (a), match each P-value to its alternative. (No explanation is needed.) Note: Even if you pick the wrong set in part (a), you can still get full credit for part (b).

9. A comparative study is performed; you are given the following information.

   - The total number of subjects equals 29.
   - The observed value of the test statistic is greater than 0.

   I used the website to obtain the exact P-value for Fisher's test for each of the three possible alternatives. These three P-values are below along with three bogus P-values.

Set 1:	0.1445	0.2890	0.9622
Set 2:	0.0762	0.1297	0.9868

   (a) Which set contains the correct P-values: 1 or 2? (No explanation is needed.)

   (b) For the set you selected in part (a), match each P-value to its alternative. (No explanation is needed.) Note: Even if you pick the wrong set in part (a), you can still get full credit for part (b).

10. 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.

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.

Chapters 5–7

15. 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.)

16. 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.)

17. Alex and Bruce each perform 200 dichotomous trials. A success is the desirable outcome; it requires more skill than does a failure. You are given the following information.

    - Each of the men achieves exactly 90 successes.
    - Alex exhibited evidence of improving skill over time; and Bruce exhibited evidence of declining skill over time.
    - Alex had successes on his first and last trials; Bruce had a success on his first trial and a failure on his last trial.
    - Alex performed better after a failure than after a success; and Bruce performed better after a success than after a failure.

For each man, identify his two tables from the tables below. Hint: For each man, choose one from Tables 1–3 and one from Tables 4–11. (Hint: If there is more than one table that satisfies the conditions stated above, just give me one of them.)

Table 1

Half	S	F	Total
1st	35	65	100
2nd	55	45	100
Total	90	110	200

Table 2

Half	S	F	Total
1st	45	55	100
2nd	45	55	100
Total	90	110	200

Table 3

Half	S	F	Total
1st	70	30	100
2nd	20	80	100
Total	90	110	200

Table 4

	Current		
Prev.	S	F	Total
S	43	46	89
F	46	64	110
Total	89	110	199

Table 5

	Current		
Prev.	S	F	Total
S	30	59	89
F	59	51	110
Total	89	110	199

Table 6

	Current		
Prev.	S	F	Total
S	43	47	90
F	47	62	109
Total	90	109	199

Table 7

	Current		
Prev.	S	F	Total
S	33	57	90
F	57	52	109
Total	90	109	199

Table 8

Prev.	Current S	Current F	Total
S	46	43	89
F	44	66	110
Total	90	109	199

Table 9

Prev.	Current S	Current F	Total
S	36	53	89
F	54	56	110
Total	90	109	199

Table 10

Prev.	Current S	Current F	Total
S	45	45	90
F	44	65	109
Total	89	110	199

Table 11

Prev.	Current S	Current F	Total
S	35	55	90
F	54	55	109
Total	89	110	199

18. Abby and Dana each perform 160 dichotomous trials. A success is the desirable outcome; it requires more skill than does a failure. You are given the following information.

- Each of the women achieves exactly 90 successes.
- Abby exhibited evidence of improving skill over time; and Dana exhibited no evidence of changing skill over time.
- Abby had failures on her first and last trials; Dana had a failure on her first trial and a success on her last trial.
- Abby performed better after a success than after a failure; and Dana performed better after a failure than after a success.

For each woman, identify her two tables from the tables below. Hint: For each woman, choose one from Tables 1–3 and one from Tables 4–11. (Hint: If there is more than one table that satisfies the conditions stated above, just give me one of them. If no table satisfies the conditions, say it is impossible.)

Table 1

Half	S	F	Total
1st	35	45	80
2nd	55	25	80
Total	90	70	160

Table 2

Half	S	F	Total
1st	50	30	80
2nd	40	40	80
Total	90	70	160

Table 3

Half	S	F	Total
1st	45	35	80
2nd	45	35	80
Total	90	70	160

Table 4

Prev.	Current S	Current F	Total
S	47	42	89
F	42	28	70
Total	89	70	159

Table 5

Prev.	Current S	Current F	Total
S	52	37	89
F	37	33	70
Total	89	70	159

Table 6

Prev.	Current S	Current F	Total
S	54	36	90
F	36	33	69
Total	90	69	159

Table 7

Prev.	Current S	Current F	Total
S	48	42	90
F	42	27	69
Total	90	69	159

Table 8

Prev.	Current S	Current F	Total
S	53	36	89
F	37	33	70
Total	90	69	159

Table 9

Prev.	Current S	Current F	Total
S	47	42	89
F	43	27	70
Total	90	69	159

Table 10

Prev.	Current S	Current F	Total
S	48	42	90
F	41	28	69
Total	89	70	159

Table 11

Prev.	Current S	Current F	Total
S	53	37	90
F	36	33	69
Total	89	70	159

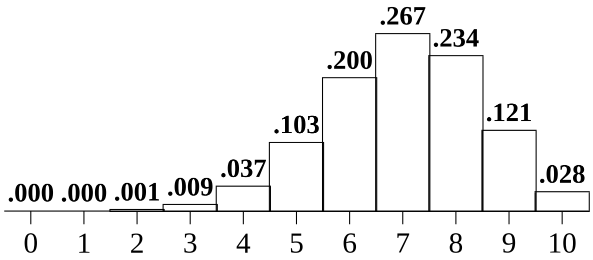
19. A box contains 14 red cards and six blue cards for a total of 20 cards. Walt is going to select $n = 10$ cards at random *with replacement* from the box. Let W denote the number of red cards that Walt obtains. Let X denote the number of blue cards that Walt obtains. Yale is going to select $n = 10$ 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.

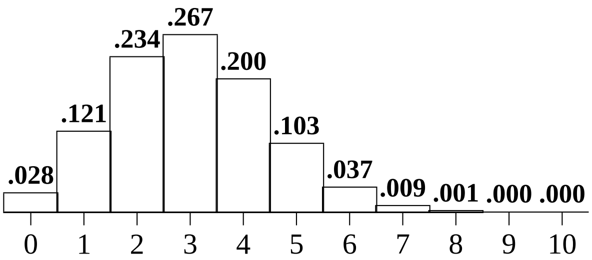
(a) Place an X next to the probability histogram of the sampling distribution of X

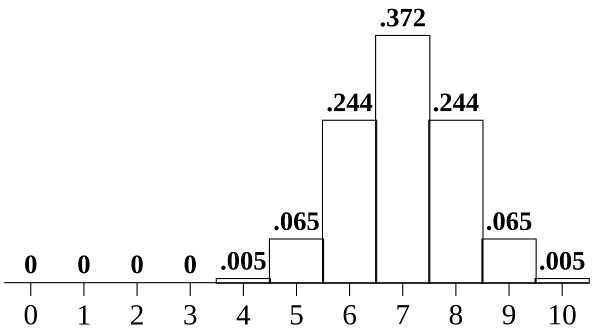
and place a Y next to the probability histogram of the sampling distribution of Y.

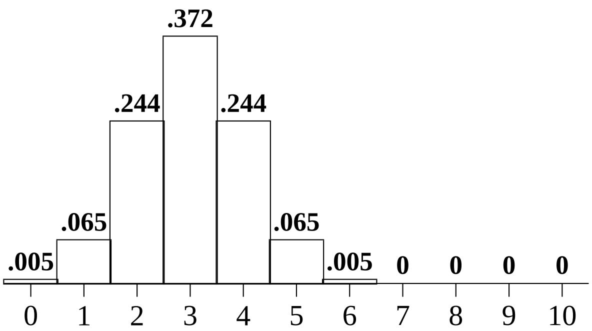
(b) What is the probability that Walt will obtain a representative sample?

(c) What is the probability that Yale will obtain a sample that is not representative because it has too many red cards?







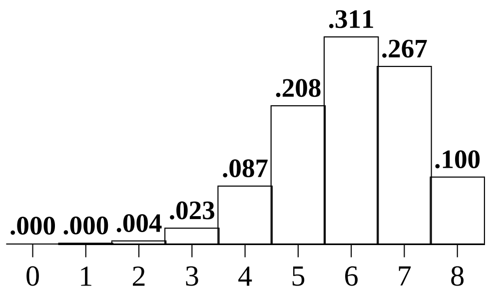


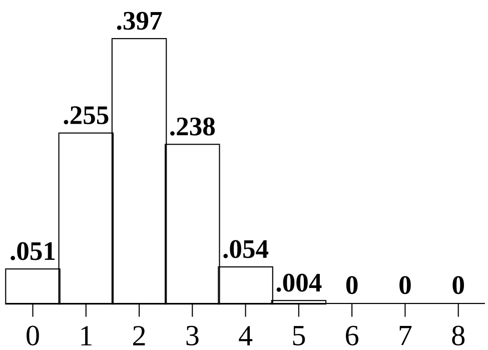
20. A box contains 15 red cards and five blue cards for a total of 20 cards. Wilma is going to select $n = 8$ cards at random *with replacement* from the box. Let W denote the number of red cards that Wilma obtains. Let X denote the number of blue cards that Wilma obtains. Yolanda is going to select $n = 8$ 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.

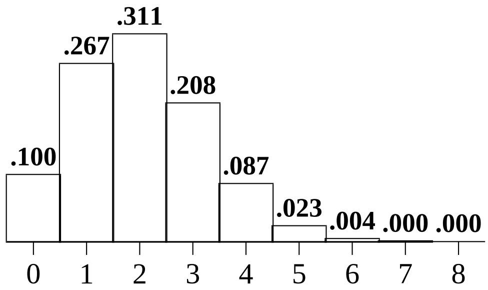
(a) Place an X next to the probability histogram of the sampling distribution of X and place a Y next to the probability histogram of the sampling distribution of Y.

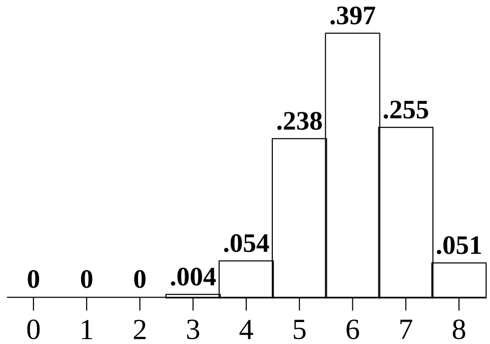
(b) What is the probability that Wilma will obtain a representative sample?

(c) What is the probability that Yolanda will obtain a sample that is not representative because it has too many red cards?









21. A random sample of size $n = 250$ yields 80 successes. Calculate the 95% confidence interval for p.

22. A random sample of size $n = 452$ yields 113 successes. Calculate the 95% confidence interval for p.

23. 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.

    (a) Calculate the point prediction of the number of ringers that George will obtain next week.

    (b) Calculate the 90% prediction interval for the number of ringers George will obtain next week.

    (c) It turns out that next week George obtains 62 ringers. Given this information, comment on your answers in parts (a) and (b).

24. Bill enjoys throwing horse shoes. Last week he tossed 140 shoes and obtained 28 ringers. (Ringers are good.) Next week he plans to throw 350 shoes. Assume that Bill's tosses satisfy the assumptions of Bernoulli trials.

    (a) Calculate the point prediction of the number of ringers that Bill will obtain next week.

    (b) Calculate the 90% prediction interval for the number of ringers Bill will obtain next week.

    (c) It turns out that next week Bill obtains 64 ringers. Given this information, comment on your answers in parts (a) and (b).

25. Bert computes a 95% confidence interval for p and obtains the interval $[0.600, 0.700]$. Note: Parts (a) and (b) are not connected: Part (b) can be answered even if one does not know how to do part (a).

    (a) Bert's boss says, "Give me a 90% confidence interval for p." Calculate the answer for Bert.

    (b) Bert's boss says, "Give me a 95% confidence interval for $p-q$." Calculate the answer for Bert. (Hint: $p-q = p-(1-p) = 2p-1$. Bert's interval says, in part, that "p is at least 0.600;" what does this tell us about $2p-1$?)

26. Maggie computes a 95% confidence interval for p and obtains the interval $[0.50, 0.75]$. Note: Parts (a) and (b) are not connected: Part (b) can be answered even if one does not know how to do part (a).

    (a) Maggie's boss says, "Give me a 95% confidence interval for p^2." Calculate the answer for Maggie. (Hint: The interval says, in part, that "p is at most 0.75;" what does this tell us about p^2?)

    (b) Maggie's boss says, "Give me a 95% confidence interval for $p-q$." Calculate the answer for Maggie. (Hint: $p-q = p-(1-p) = 2p-1$. The interval says, in part, that "p is at most 0.75;" what does this tell us about $2p-1$?)

27. Bob selects independent random samples from two populations and obtains the values $\hat{p}_1 = 0.700$ and $\hat{p}_2 = 0.500$. He constructs the 95% confidence interval for $p_1 - p_2$ and gets:

$$0.200 \pm 1.96(0.048) = 0.200 \pm 0.094.$$

Note that 0.048 is called the estimated standard error of $\hat{p}_1 - \hat{p}_2$ (the ESE of the estimate).

Tom wants to estimate the mean of the success rates:

$$\frac{p_1 + p_2}{2}.$$

(a) Calculate Tom's point estimate.

(b) Given that the estimated standard error of $(p_1 + p_2)/2$ is 0.024, calculate the 95% confidence interval estimate of $(p_1 + p_2)/2$. Hint: The answer has our usual form:

$$\text{Pt. est. } \pm 1.96 \times \text{ESE of the estimate.}$$

28. 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. You do **not** need to explain your reasoning.

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

Group	S	F	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	S	F	Tot	Gp	S	F	Tot
1	3	17	20	1	40	40	80
2			40	2			60
Tot			60	Tot			140

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

Group	S	F	Total
1	43	57	100
2	36	64	100
Total	79	121	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	3	17	20	1	40	40	80
2			40	2			60
Tot			60	Tot			140

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

Group	S	F	Total
1	45	55	100
2	38	62	100
Total	83	117	200

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

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	5	15	20	1	40	40	80
2			60	2			40
Tot			80	Tot			120

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

Group	S	F	Total
1	44	56	100
2	43	57	100
Total	87	113	200

Complete these tables so that Simpson's Paradox **does not** occur (see Course Notes). Note that there might be more than one possible correct answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	4	16	20	1	40	40	80
2			60	2			40
Tot			80	Tot			120

Chapters 8, 15, 16, and 13

33. Below is the table of population counts for a condition and its screening test. (Recall that A means the condition is present and B means the screening test is positive.) On parts (a)–(e) below, show enough work for me to understand how you obtained your answer; e.g. don't just write down "0.5."

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

(a) What proportion of the population is free of the condition?

(b) What proportion of the population has the condition and would test positive?

(c) Of those who have the condition, 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?

34. Below is the table of population counts for a condition and its screening test. (Recall that A means the condition is present and B means the screening test is positive.) On parts (a)–(f) below, report your answers to three digits of precision, for example 0.194.

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

(a) What proportion of the population is free of the condition?

(b) What proportion of the population has the condition and would test positive?

(c) Of those who have the condition, 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?

35. Consider all courtroom trials with a single defendant who is charged with a felony. Suppose that you are given the following probabilities for this situation.

Eighty-two percent of the defendants are, in fact, guilty. Given that the defendant is guilty, there is a 75 percent chance the jury will convict the person. Given that the defendant is not guilty, there is a 40 percent chance the jury will convict the person.

For simplicity, assume that the only options available to the jury are: to convict or to release the defendant.

(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?

36. Consider all courtroom trials with a single defendant who is charged with a felony. Suppose that you are given the following probabilities for this situation.

    Seventy-five percent of the defendants are, in fact, guilty. Given that the defendant is guilty, there is a 70 percent chance the jury will convict the person. Given that the defendant is not guilty, there is a 40 percent chance the jury will convict the person.

    For simplicity, assume that the only options available to the jury are: to convict or to release the defendant.

    (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?

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 four 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.

    $$[24, 41], [30, 39], [20, 33], \text{ and } [35, 45].$$

    (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 median? (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 median? (Hint: The answer is not any of the four confidence intervals.)

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

38. 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 median of the population. The intervals are below.

    $$[14, 31], [20, 29], [10, 23], \text{ and } [25, 35].$$

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

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

39. Homer performs three simulation studies. His population is skewed to the right. For one study he has his computer generate 10,000 random samples of size $n = 10$ from the population. For each random sample, the computer calculates the Gosset 95% confidence interval for μ and checks to see whether the interval is correct. His second study is like his first, but $n = 100$. Finally, his third study is like the first, but $n = 200$. In one of his studies, Homer obtains 9,504 correct intervals; in another he obtains 9,478 correct intervals; and in the remaining study he obtains 8,688 correct intervals.

    Based on what we learned in class, match each sample size to its number of correct intervals. Explain your answer.

40. 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.

41. 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.

42. 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

(a) Construct the 95% confidence interval for $\mu_X - \mu_Y$.

(b) Obtain the P-value for the alternative $\mu_X \neq \mu_Y$. Show your work. You will receive **no credit** for simply reporting your answer.

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

Pop.	Mean	Stand. Dev.	Sample size
1	73.00	10.00	14
2	62.50	6.00	8

(a) Construct the 95% confidence interval for $\mu_X - \mu_Y$.

(b) Obtain the P-value for the alternative $\mu_X \neq \mu_Y$.

44. A regression analysis yields the line

$$\hat{y} = 32 + 0.4x.$$

One of the subjects, Racheal, has $x = 60$ and $y = 52$.

(a) Calculate Racheal's predicted value, $\hat{y}$.

(b) Calculate Racheal's residual.

45. A regression analysis yields the line

$$\hat{y} = 18 + 0.25x.$$

One of the subjects, Mary, has $x = 40$ and $y = 32$.

(a) Calculate Mary's predicted value, $\hat{y}$.

(b) Calculate Mary's residual.

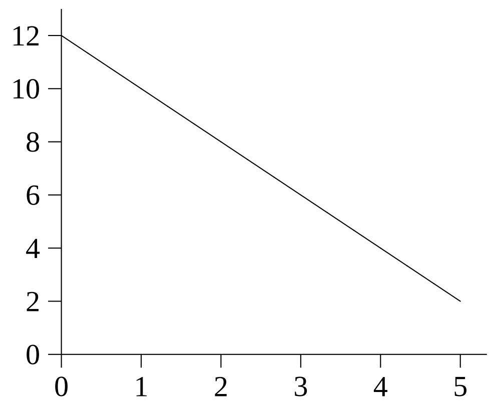
46. Fifty students take midterm and final exams. On the midterm exam, the mean score is 45.0 and the standard deviation is 7.00. On the final exam, the mean score is 85.0 with a standard deviation of 14.00. The correlation coefficient of the two scores is 0.64.

Obtain the least squares regression line for using the final exam score to predict the midterm exam score.

47. 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.

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



(a) Locate the point that has $x = 3$ and $y = 8$; put an A at that point.

(b) Locate the point that has $x = 5$ and $e = 2$; put a B at that point.

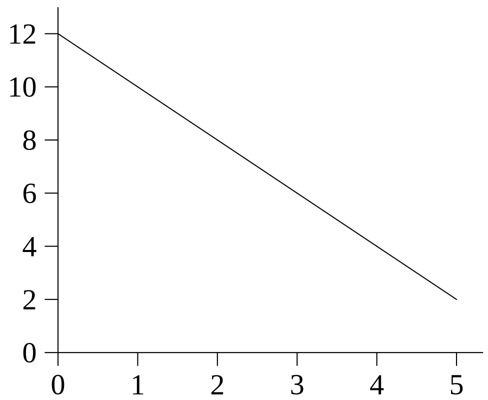
(c) Locate the point that has $y = 6$ and $e = -2$; put a C at that point.

(d) Locate the point that has $\hat{y} = 6$ and $e = -4$; put a D at that point.

(e) Draw the line that represents all points for which $e = -2$.

(f) Given that $\bar{x} = 3$, what is the value of $\bar{y}$?

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



(a) Locate the point that has $x = 2$ and $y = 6$; put an A at that point.

(b) Locate the point that has $x = 4$ and $e = 4$; put a B at that point.

(c) Locate the point that has $y = 8$ and $e = -2$; put a C at that point.

(d) Draw the line that represents all points for which $e = -3$.

(e) Given that $\bar{x} = 4$, what is the value of $\bar{y}$?

50. 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.

51. 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

Chapters 1, 12, 2, and 3

1. (a) The median is between 9.00 and 9.50; here is why. There are 100 observations in the data set. Thus, the median is the average of the numbers in positions 50 and 51 of the sorted data. Look at the picture from left to right. The sum of the frequencies of the first four rectangles is 40. Add the fifth rectangle and the sum of the frequencies is 56. Thus, the 50th and 51st sorted values are in the fifth rectangle. The boundaries of the fifth rectangle are 9.00 and 9.50.

   (b) The first histogram is bell-shaped; thus, the empirical rule should work well; i.e. approximately 68% of the data should lie in the interval $\bar{x} \pm s$. For $s = 1.5$ it contains 54% of the data; for $s = 2.0$ it contains 68% of the data; and for $s = 2.5$ it contains 78% of the data. Thus, $s = 2.0$ is the most reasonable answer.

   (c) The second histogram is strongly skewed to the right; thus, the empirical should not work very well. In fact (and I mentioned this more than once during class), for a strongly skewed distribution much more than 68% of the data will lie in the interval $\bar{x} \pm s$. For $s = 1.0$ it contains 38% of the data; for $s = 1.5$ it contains 64% of the data; and for $s = 2.0$ it contains 81% of the data. Thus, $s = 2.0$ is the most reasonable answer.

2. (a) The key idea is that the concept of the mean is tied to the concept of the total. If you know one of these, you can calculate the other. Thus, Al's total is 30 and Bev's total is 90. Thus, Carol's total is $30 + 90 = 120$ and her mean is $120/8 = 15$.

   (b) There are many possible answers; here is one.

   Al: 10, 10, 10
   Bev: 18, 18, 18, 18, 18

   Combine these and the median is 18.

   Al: 10, 10, 10
   Bev: 10, 10, 18, 26, 26

   Combine these and the median is 10.

3. (a) The range is $88.9 - 14.1 = 74.8$. 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 and upper halves of the data both have 20 observations. The first quartile is the mean of 55.4 and 57.6; i.e. 56.5. The third quartile is the mean of 72.4 and 72.9; i.e. 72.65. The IQR is 16.15.

   (b) 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 $32/40 = 0.80$.

   (c) It is larger than the predicted 68%.

   (d) The range is $88.9 - 46.0 = 42.9$. The median is the number in position 20; i.e. 64.3. The lower and upper halves of the data both have 19 observations. The first quartile is 57.6; the third quartile is 72.9. Thus, the IQR is 15.3.

   (e) From (b), the mean of the 40 observations is 63.50. Thus, the total of the 40 observations is $40(63.50) = 2540$. Therefore, the total of the 39 remaining observations is $2540 - 14.1 = 2525.9$ and the mean is $2525.9/39 = 64.77$.

4. (a) The first thing to do is calculate x:

   $$x = \hat{p}_1 - \hat{p}_2 = 6/18 - 2/12 = 0.1667.$$

   Thus, the P-value is

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

(b) The P-value is

$$P(X \leq -0.2500) + P(X \geq 0.2500) =$$
$$0.1371 + P(X \geq 0.3056) =$$
$$0.1371 + 0.0727 = 0.2098.$$

(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 $x = -0.3889$.

5. (a) The first thing to do is calculate x:

$$x = \hat{p}_1 - \hat{p}_2 = 6/22 - 2/16 = 0.1477.$$

Thus, the P-value is

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

(b) The P-value is

$$P(X \leq -0.1761) + P(X \geq 0.1761) =$$
$$0.1808 + P(X \geq 0.2557) =$$
$$0.1808 + 0.0623 = 0.2431.$$

(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 $x = -0.2841$.

6. (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.

7. (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.

8. (a) Set 1. Here is why. The issue is symmetry. B/c $n = 33$ is an odd number, we know that the study is not balanced and we know that $m_1 \neq m_2$. Thus, the sampling distribution is not symmetric. As a result, the P-value for $\neq$ is not twice the smaller of the other two P-values. Thus, set 2 cannot be the P-values b/c 0.2890 is twice 0.1445.

(b) First, let P_1 be the P-value for the first alternative, $>$; let P_2 be the P-value for the second alternative, $<$; and let P_3 be the P-value for the third alternative, $\neq$.

We have the following facts.

- $P_1 + P_2 > 1$.
- $P_3 \geq P_1$ (b/c $x > 0$; if $x < 0$, then $P_3 \geq P_2$.)

From the first fact, either P_1 or P_2 must equal 0.9233. If, however, $P_1 = 0.9233$, then the second fact would be violated. Thus, $P_2 = 0.9233$. Now, from the second fact, $P_1 = 0.2450$ and $P_3 = 0.4688$.

BTW, if you chose Set 2 in part (a) you can get full credit for part (b) with the following answers (same reasoning as above). $P_2 = 0.9625$, $P_1 = 0.1445$ and $P_3 = 0.2890$.

9. (a) Set 2. One of the P-values for set 1 is twice as large as another. As discussed in class, this implies that the sampling distribution is symmetric. With an odd sample size,

however, the sampling distribution cannot be symmetric.

(b) The P-value for $>$ is 0.0762; for $<$ it is 0.9868; and for $\neq$ it is 0.1297. BTW, if you answered set 1 for (a), you received full credit on (b) for the following answers: 0.1445 for $>$; 0.9622 for $<$; and 0.2890 for $\neq$.

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

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

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

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

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

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

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.

Chapters 5–7

15. (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. $P(SFFS) =$

$$0.3234(0.6766)(0.6766)(0.3234) =$$

$$0.0479.$$

16. (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. $P(SFFF) =$

$$0.2739(0.7261)(0.7261)(0.7261) =$$

$$0.1049.$$

17. Tables 1 and 5 are Alex's; Tables 3 and 10 are Bruce's.

18. Tables 1 and 6 are Abby's; Tables 3 and 9 are Dana's.

19. (a) The second picture is for X and the third picture is for Y.

    (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,

$$P(Y > 7) = 0.244 + 0.065 + 0.005 =$$

$$0.314.$$

20. (a) The third picture is for X and the fourth picture is for Y.

    (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,

$$P(Y > 6) = 0.255 + 0.051 = 0.306.$$

21. First, $\hat{p} = 80/250 = 0.320$. The 95% confidence interval is

$$0.320 \pm 1.96\sqrt{\frac{0.32(0.68)}{250}} =$$

$$0.320 \pm 0.058 = [0.262, 0.378].$$

22. First, $\hat{p} = 113/452 = 0.250$. The 95% confidence interval is

$$0.250 \pm 1.96\sqrt{\frac{0.25(0.75)}{452}} =$$

$$0.250 \pm 0.040 = [0.210, 0.290].$$

23. (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.

24. (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 ringers, the point prediction is too large by 6, but the prediction interval is correct b/c 64 is between 47 and 93.

25. (a) The half-width of the interval is 0.050. Also, it is

$$1.96\sqrt{\hat{p}\hat{q}/n}.$$

    (You could plug in the values of $\hat{p}$ and $\hat{q}$ and compute n, but you don't need to do so.) Thus,

$$\sqrt{\hat{p}\hat{q}/n} = 0.050/1.96,$$

    and

$$1.645\sqrt{\hat{p}\hat{q}/n} = [1.645(0.050]/1.96 = 0.042.$$

    Thus, the 90% confidence interval is

$$0.650 \pm 0.042 = [0.608, 0.692].$$

(b) The confidence interval states

$$0.600 \leq p \leq 0.700.$$

This yields the following inequalities.

$$1.200 \leq 2p \leq 1.400, \text{ and}$$

$$0.200 \leq 2p - 1 \leq 0.400.$$

Thus, $[0.200, 0.400]$ is the 95% confidence interval for $p - q$.

26. (a) The CI states that $p \leq 0.75$; thus, $p^2 \leq (0.75)^2 = 0.5625$. Similarly, $p^2 \geq 0.2500$. Thus, the 95% CI for p^2 is $[0.2500, 0.5625]$.

(b) The confidence interval states

$$0.50 \leq p \leq 0.75.$$

This yields the following inequalities.

$$1.00 \leq 2p \leq 1.50, \text{ and}$$

$$0.00 \leq 2p - 1 \leq 0.50.$$

Thus, $[0.00, 0.50]$ is the 95% confidence interval for $p - q$.

27. (a) The point estimate is

$$\frac{\hat{p}_1 + \hat{p}_2}{2} = \frac{0.700 + 0.500}{2} = 0.600.$$

(b) The 95% confidence interval is

$$0.600 \pm 1.96(0.024) = 0.600 \pm 0.047 =$$

$$[0.553, 0.647].$$

28. 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%.

29. 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.

30. 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 36 in the collapsed table. This combination of restrictions is incompatible.

31. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need $c \geq 16$ in Subgroup A and $c \geq 21$ in Subgroup B. Also, the two c's must sum to the 38 in the collapsed table. There are two possible answers: 16 and 22; 17 and 21.

32. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To fail to get a reversal in both tables, we need $c \leq 12$ in Subgroup A or $c \leq 20$ in Subgroup B. Also, the two c's must sum to the 43 in the collapsed table. Many students made the error of thinking this combination is impossible, but its not. Note the word 'or.' We don't need to fail to reverse in both tables; just in one. Thus, there are several answers that work: 12 and 31; 11 and 32; …; 3 and 40; 23 and 20; 24 and 19; …; 43 and 0.

Chapters 8, 15, 16, and 13

33. (a) $740/860 = 0.860$.

(b) $108/860 = 0.126$.

(c) $12/120 = 0.100$.

(d) The number of correct results is $108 + 698 = 806$. Thus, the proportion is $806/860 = 0.937$.

(e) The number of incorrect results is $12 + 42 = 54$. Thus, the proportion is $12/54 = 0.222$.

34. (a) $612/720 = 0.850$.

(b) $96/720 = 0.133$.

(c) $12/108 = 0.111$.

(d) The number of correct results is $96 + 564 = 660$. Thus, the proportion is $660/720 = 0.917$.

(e) The number of incorrect results is $12 + 48 = 60$. Thus, the proportion is $12/60 = 0.200$.

35. (a) Let A denote that the defendant is guilty and let B denote conviction. First, use the information to obtain the following table.

	B	B^c	Total
A	0.615	0.205	0.820
A^c	0.072	0.108	0.180
Total	0.687	0.313	1.000

The proportion of the defendants who will be convicted is 0.687.

(b) From the definition,

$$P(A|B) = P(AB)/P(B) =$$

$$0.615/0.687 = 0.895.$$

(c) $0.615 + 0.108 = 0.723$.

(d) The probability of an incorrect decision is $0.205 + 0.072 = 0.277$. Thus, the conditional probability is $0.205/0.277 = 0.740$.

36. (a) Let A denote that the defendant is guilty and let B denote conviction. First, use the information to obtain the following table.

	B	B^c	Total
A	0.525	0.225	0.750
A^c	0.100	0.150	0.250
Total	0.625	0.375	1.000

The proportion of the defendants who will be convicted is 0.625.

(b) From the definition,

$$P(A|B) = P(AB)/P(B) =$$

$$0.525/0.625 = 0.840.$$

(c) $0.525 + 0.150 = 0.675$.

(d) The probability of an incorrect decision is $0.225 + 0.100 = 0.325$. Thus, the conditional probability is $0.225/0.325 = 0.692$.

37. (a) $39 < \nu \leq 41$. Note: Throughout this problem you will receive full credit even if you confuse '$<$' with '$\leq$.' You will, however, lose credit for writing $\nu = 40$ or 41; i.e. if you assume that ν must be an integer.

(b) $33 < \nu < 35$.

(c) The maximum number of correct CIs is 3. It will occur if $30 \leq \nu \leq 33$ or $35 \leq \nu \leq 39$.

38. (a) $14 \leq \nu < 20$. Note: Throughout this problem you will receive full credit even if you confuse '$<$' with '$\leq$.'

(b) $23 < \nu < 25$.

(c) The maximum number of correct CIs is 3. It will occur if $20 \leq \nu \leq 23$ or $25 \leq \nu \leq 29$.

39. We know that sometimes Gosset does not work well, especially for a skewed population. But if Gosset performs poorly for a particular n, it will do better if n is increased. In this example, one performance is poor—the one with 8,688 correct CIs. This must be for $n = 10$. If Gosset performs well for a specific n, it will also perform well for any larger n. The 9,478 and 9,504 correct are both good performances; thus, we cannot tell which goes with $n = 100$ and which goes with $n = 200$.

40. (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 $[530, 671]$.

41. (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 $[56.0, 62.5]$.

42. (a) First,

$$s_p^2 = \frac{16(100) + 9(36)}{16 + 9} = 76.96,$$

and $s_p = 8.773$. The 95% CI is

$$8.00 \pm 2.060(8.773)\sqrt{\frac{1}{17} + \frac{1}{10}} =$$

$$8.00 \pm 2.060(3.496) = 8.00 \pm 7.20 =$$

$$[0.80, 15.20].$$

(b) The observed value of the test statistic is

$$t = \frac{8.00}{3.496} = 2.288,$$

with 25 degrees of freedom. The P-value is between 0.02 and 0.05.

43. (a) First,

$$s_p^2 = \frac{13(100) + 7(36)}{13 + 7} = 77.6,$$

and $s_p = 8.809$. The 95% CI is

$$10.50 \pm 2.086(8.809)\sqrt{\frac{1}{14} + \frac{1}{8}} =$$

$$10.50 \pm 2.086(3.904) = 10.50 \pm 8.14 =$$

$$[2.36, 18.64].$$

(b) The observed value of the test statistic is

$$t = \frac{10.50}{3.904} = 2.690,$$

with 20 degrees of freedom. The P-value is between 0.01 and 0.02.

44. (a) Her predicted value is

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

(b) Her residual is

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

45. (a) Her predicted value is

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

(b) Her residual is

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

46. 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

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

47. 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,

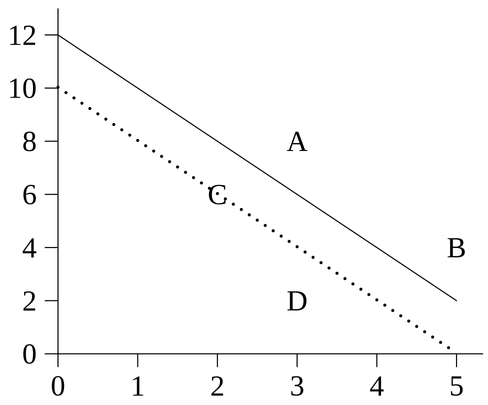
$$b_1 = 0.70(7/10) = 0.49, \text{ and}$$

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

The regression line is

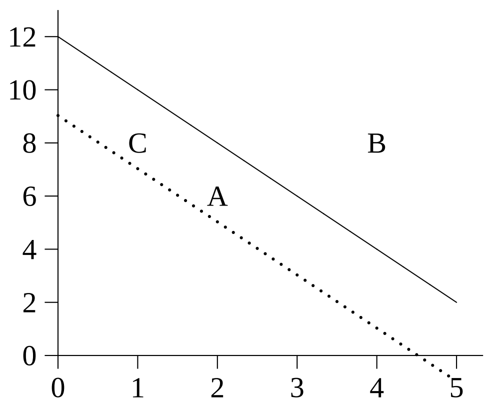
$$\hat{y} = 38.05 + 0.49x.$$

48. The points are plotted below.



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

49. The points are plotted below.



Recall that if $x = \bar{x}$, then $\hat{y} = \bar{y}$. Thus, $\bar{y} = 12 - 2(4) = 4$.

50. 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

$$\hat{y} = b_0 + b_1 x.$$

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

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

51. 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

$$\hat{y} = b_0 + b_1 x.$$

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

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