Extra Homework; Statistics 301;
Professor Wardrop
Fall 2007

Section 2.3

1. Consider a balanced study with eight subjects, identified as A, B, C, D, E, G, H, and J. In the actual study,

   - A, B, C and D are assigned to the first treatment, and
   - There are exactly four successes, and they are obtained by A, B, C, and H.

   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 A, D, E, and G 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 incorrect about subjects C, D, H, and J, but correct about subjects A, B, E, and G. This means that for subjects C, D, H, and J, his/her/its response will change if the treatment changes.

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

2. Consider a unbalanced study with nine subjects, identified as A, B, C, D, E, G, H, J, and K. In the actual study,

   - A, B, C, D, and E are assigned to the first treatment, and
   - There are exactly five successes, and they are obtained by B, C, E, H, and J.

   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 A, C, D, G, and K 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 incorrect about subjects C, D, E, G, H, J, and K, but correct about subjects A, and B. This means that for subjects C, D, E, G, H, J, and K, his/her/its response will change if the treatment changes.

   For the assignment that puts A, E, G, H, and J on the first treatment, and the other subjects on the second treatment, determine the response for each of the nine subjects.

3. An unbalanced CRD yields the data below.

Treatment	S	F	Total
1	a	b	10
2	c	d	5
Total	10	5	15

   On the assumption the Skeptic is correct, list all possible values of the test statistic.

4. A CRD yields the following data.

Treatment	S	F	Total
1	a	b	5
2	c	d	4
Total	6	3	9

On the assumption the Skeptic is correct, determine all possible values of the test statistic.

Section 2.5

5. 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	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.30	0.0003	0.0003	1.0000
-0.25	0.0029	0.0032	0.9997
-0.20	0.0151	0.0184	0.9968
-0.15	0.0514	0.0697	0.9816
-0.10	0.1193	0.1890	0.9303
-0.05	0.1957	0.3848	0.8110
0.00	0.2305	0.6152	0.6152
0.05	0.1957	0.8110	0.3848
0.10	0.1193	0.9303	0.1890
0.15	0.0514	0.9816	0.0697
0.20	0.0151	0.9968	0.0184
0.25	0.0029	0.9997	0.0032
0.30	0.0003	1.0000	0.0003

(a) Find the P-value for the second alternative ($p_1 < p_2$) if $x = -0.15$.

(b) Determine the P-value for the third alternative ($p_1 \neq p_2$) if $x = -0.25$.

(c) Determine the value of x and the P-value that satisfy the following condition: The data are statistically significant but not highly statistically significant for the first alternative ($p_1 > p_2$).

6. Pam 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	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.289	0.0002	0.0002	1.0000
-0.244	0.0016	0.0018	0.9998
-0.200	0.0089	0.0107	0.9982
-0.156	0.0330	0.0437	0.9893
-0.111	0.0851	0.1288	0.9563
-0.067	0.1576	0.2864	0.8712
-0.022	0.2136	0.5000	0.7136
0.022	0.2136	0.7136	0.5000
0.067	0.1576	0.8712	0.2864
0.111	0.0851	0.9563	0.1288
0.156	0.0330	0.9893	0.0437
0.200	0.0089	0.9982	0.0107
0.244	0.0016	0.9998	0.0018
0.289	0.0002	1.0000	0.0002

(a) Find the P-value for the second alternative ($p_1 < p_2$) if $x = -0.111$.

(b) Determine the P-value for the first alternative if $x = 0.289$.

(c) Determine all values of x and the P-value that satisfy the following condition. The data are statistically significant but not highly statistically significant for the third alternative ($p_1 \neq p_2$).

7. Two comparative studies with dichotomous responses, randomization, and two treatments are performed. The first study has $n = 35$ and the second study is balanced.

The observed value of the test statistic, x, is a positive number for the first study and a negative number for the second study. The P-values for all three alternatives are obtained for both studies. The six (sorted) P-values are below.

0.0433, 0.0866, 0.3297, 0.5060, 0.8700, and 0.9865.

Match each P-value with its study and alternative.

	Alternative		
Study	$>$	$<$	$\neq$
First			
Second			

8. Two comparative studies, with two treatments, dichotomous responses and randomization are performed. The first study is balanced.

   The observed value of the test statistic, x, is a positive number for the first study and a negative number for the second study. The P-values for all three alternatives are obtained for both studies. The six (sorted) P-values are below.

   0.1001, 0.1223, 0.2002, 0.2187,
   0.9432, and 0.9565.

   Match each P-value with its study and alternative. **Hint: For the second study, for the actual x,**

$$P(X = x) = 0.0655.$$

	Alternative		
Study	$>$	$<$	$\neq$
First			
Second			

Section 2.4

9. **(Extra Credit.)** A balanced CRD with $n = 60$ 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.)** A balanced CRD with $n = 80$ subjects is performed. The study yields a total of 30 successes. Write an expression for $P(X = 0.25)$ using binomial coefficients; do not compute the answer.

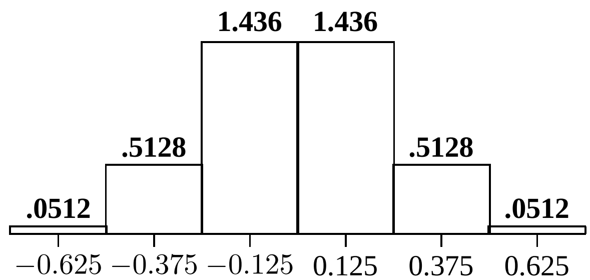
Section 3.1

11. Below is the sampling distribution of the test statistic for Fisher's Test for an unbalanced CRD.

x	$P(X = x)$	x	$P(X = x)$
-0.8	0.0070	0.1	0.3916
-0.5	0.0932	0.4	0.1632
-0.2	0.3263	0.7	0.0186

   Draw the probability histogram for this distribution.

12. Below is the probability histogram for the test statistic for a Fisher's test. Obtain the sampling distribution and present it in a table (i.e. a column of x's and a column of probabilities).



13. Maria draws a probability histogram for a Fisher's test. You are given the following facts:

   - The study is balanced.
   - There is a rectangle centered at 0.000 and a rectangle centered at 0.025. These rectangles "touch;" that is, there are no other rectangles between them.
   - The rectangle centered at 0.000 has a height of 5.784.

   (a) Calculate $P(X = 0.000)$.

   (b) Given that the actual value of the test statistic is $x = 0.025$, calculate the exact P-value for the alternative $p_1 > p_2$.

14. Nancy draws a probability histogram for a Fisher's test. You are given the following facts:

   - The study is balanced.
   - There is a rectangle centered at 0.00 and a rectangle centered at 0.08. These rectangles "touch;" that is, there are no other rectangles between them.
   - The rectangle centered at 0.00 has a height of 2.835.

   (a) Calculate $P(X = 0)$.

   (b) Given that the actual value of the test statistic is $x = 0.08$, calculate the P-value for the alternative $p_1 > p_2$.

Section 3.2

15. Below is the sampling distribution of the test statistic for Fisher's Test for an unbalanced CRD.

x	$P(X=x)$	x	$P(X=x)$
-0.8	0.0070	0.1	0.3916
-0.5	0.0932	0.4	0.1632
-0.2	0.3263	0.7	0.0186

I performed a simulation experiment with 5,350 runs. The frequencies of occurrences of the six values of the test statistic were obtained and then sorted:

34, 107, 482, 868, 1769, and 2090.

Which of these five numbers is the observed frequency of $x = 0.4$? Briefly justify your answer.

16. Below is the sampling distribution of the test statistic for Fisher's Test for an unbalanced CRD.

x	$P(X=x)$	x	$P(X=x)$
-0.8	0.0037	0.1	0.4396
-0.5	0.0732	0.4	0.1538
-0.2	0.3297		

I performed a simulation experiment with 6,250 runs. The frequencies of occurrence of the five values of the test statistic were obtained and then sorted:

17, 450, 975, 2073, and 2735.

Which of these five numbers is the observed frequency of $x = 0.4$? Briefly justify your answer.

17. 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:

1973, 2021, and 6006.

(a) What is your approximation of $P(B)$?

(b) What is the best approximation of $P(C)$?

Section 3.4

18. An unbalanced CRD has a total of 250 subjects, with 100 subjects on treatment 1. The total number of successes is 135, with 45 of the successes on the first treatment.

Use the standard normal curve to obtain an approximate P-value for Fisher's test with the third alternative ($\neq$).

19. An unbalanced CRD has a total of 280 subjects, with 100 subjects on treatment 1. The total number of successes is 153, with 45 of the successes on the first treatment.

Use the standard normal curve to obtain an approximate P-value for Fisher's test with the third alternative ($\neq$).

20. An unbalanced CRD has a total of 380 subjects, with 80 subjects on treatment 1. The total number of successes is 123, with 24 of the successes on the first treatment.

Use the standard normal curve to obtain an approximate P-value for Fisher's test with the second alternative ($<$).

Section 5.2

21. A box contains 10 red cards and five blue cards for a total of 15 cards. Walter is going to select $n = 9$ cards at random *with replacement* from the box. Let W denote the number of red cards that Walter obtains. Let X denote the number of blue cards that Walter obtains. Yvonne is going to select $n = 9$ cards at random *without replacement* from the box. Let Y denote the number of red cards that Yvonne obtains. Finally, let Z denote the number of blue cards that Yvonne 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.

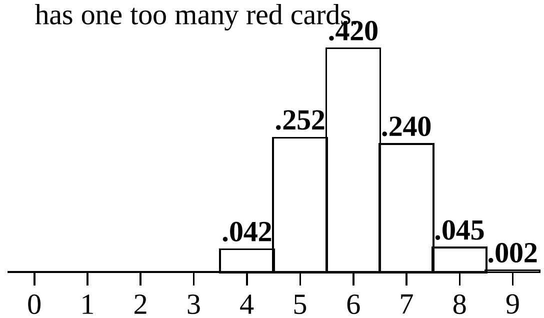
(a) Match each probability histogram with its variable (W, X, Y, and Z).

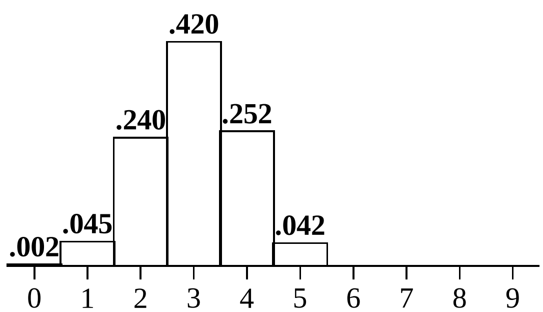
(b) Use the appropriate probability histogram to determine the probability that Yvonne will obtain exactly two blue cards.

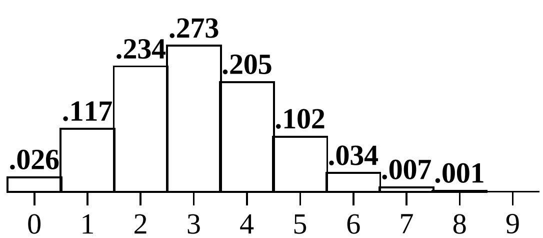
(c) Use the appropriate probability histogram to determine the probability that Yvonne will obtain two or fewer blue cards.

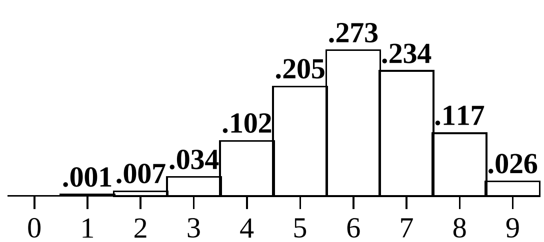
(d) Let y represent the probability that Yvonne obtains a representative sample. What is the the numerical value of y?

(e) Compute the probability that Walter obtains a sample that is either representative or misses being representative because it has one too many red cards.









22. A box contains 10 red cards and six blue cards for a total of 16 cards. Wilbur is going to select $n = 8$ cards at random *with replacement* from the box. Let W denote the number of red cards that Wilbur obtains. Let X denote the number of blue cards that Wilbur 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 below. The number above each rectangle is its height which also equals its area.

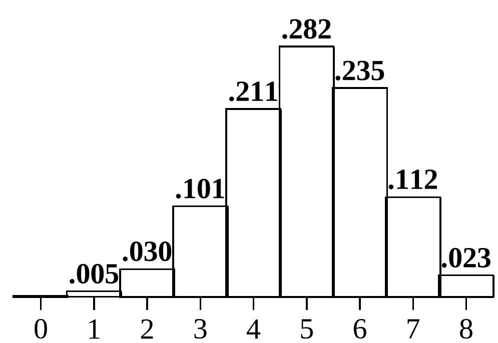
(a) Match each probability histogram with its variable (W, X, Y, and Z).

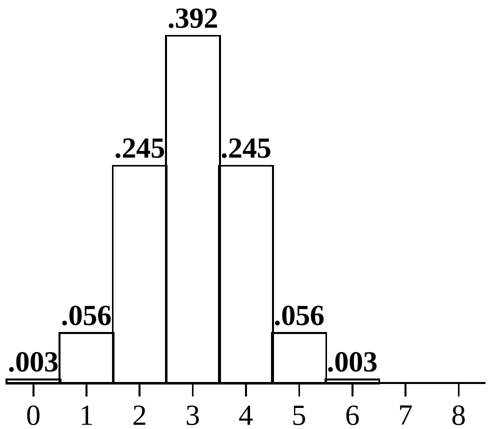
(b) Use the appropriate probability histogram to determine the probability that Yolanda will obtain exactly six red cards.

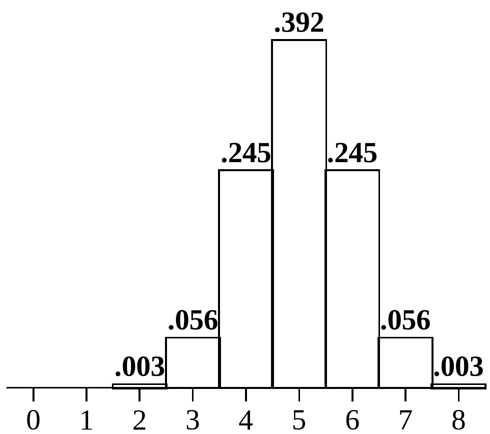
(c) Use the appropriate probability histogram to determine the probability that Yolanda will obtain six or more red cards.

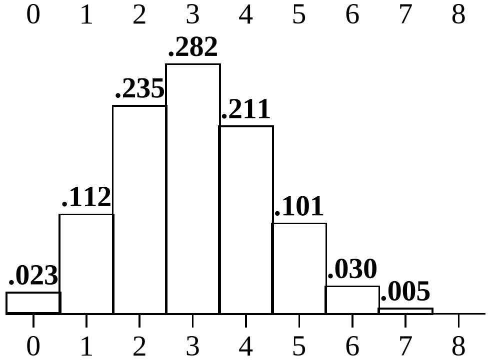
(d) Let w represent the probability that Wilbur obtains a representative sample. What is the the numerical value of w?

(e) Compute the probability that Yolanda obtains a sample that is either representative or misses being representative because it has one too many red cards.









Section 5.3

23. Carol performs 100 dichotomous trials and obtains the following data.

Half	S	F	Total
First	a	b	50
Second	c	d	50
Total	34	66	100

For each statement below, determine *any one* value of c that makes the statement true for the data above. Be careful. Note that I am asking for the value of c. Also, note that for some statements, there might be more than one possible value of c that works.

A. There is evidence that Carol's ability improved during the course of the study.
B. There is no evidence that Carol's ability changed during the course of the study.
C. There is evidence that Carol's ability declined during the course of the study.

24. Barb performs 201 dichotomous trials and obtains the following data.

	Current		
Previous	S	F	Total
S	a	b	40
F	c	d	160
Total	40	160	200

For each statement below, determine *any one* value of c that makes the statement true for the data above. Be careful. Note that I am asking for the value of c. Also, note that for some statements, there might be more than one possible value of c that works.

A. There is no evidence that the outcome of the previous trial has any influence on the outcome of the current trial.
B. There is evidence that Barb performs better after a success than after a failure.
C. There is evidence that Barb performs better after a failure than after a success.

25. Alice, Betty, and Carla each perform 100 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 forty successes.
- Alice exhibited evidence of improving skill over time; Carla exhibited evidence of declining skill over time; and Betty exhibited no evidence of changing skill over time.

- Alice performed much better after a success than after a failure; Betty performed much better after a failure than after a success; and Carla performed nearly the same after a failure as after a success.
- Alice had successes on her first and last trials; Betty had failures on her first and last trials; and Carla had a success on her first trial and a failure on her last trial.

For each woman, identify her two tables from the tables below. Note: For each woman, select one from Tables 1–4, and one from Tables 5–12.

Table 1

Half	S	F	Tot.
1st	25	25	50
2nd	15	35	50
Total	40	60	100

Table 2

Half	S	F	Tot.
1st	15	25	40
2nd	35	25	60
Total	50	50	100

Table 3

Half	S	F	Tot.
1st	20	30	50
2nd	20	30	50
Total	40	60	100

Table 4

Half	S	F	Tot.
1st	15	35	50
2nd	25	25	50
Total	40	60	100

Table 5

	Current		
Prev.	S	F	Tot.
S	5	35	40
F	34	25	59
Total	39	60	99

Table 6

	Current		
Prev.	S	F	Tot.
S	16	24	40
F	23	36	59
Total	39	60	99

Table 7

	Current		
Prev.	S	F	Tot.
S	19	20	39
F	20	40	60
Total	39	60	99

Table 8

	Current		
Prev.	S	F	Tot.
S	10	29	39
F	29	31	60
Total	39	60	99

Table 9

	Current		
Prev.	S	F	Tot.
S	23	16	39
F	17	43	60
Total	40	59	99

Table 10

	Current		
Prev.	S	F	Tot.
S	16	23	39
F	24	36	60
Total	40	59	99

Table 11

	Current		
Prev.	S	F	Tot.
S	25	15	40
F	15	44	59
Total	40	59	99

Table 12

	Current		
Prev.	S	F	Tot.
S	10	30	40
F	30	29	59
Total	40	59	99

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

Table 1

Half	S	F	Total
1st	30	50	80
2nd	30	50	80

Table 2

Half	S	F	Total
1st	60	100	160
2nd	40	120	160

Table 3

Half	S	F	Total
1st	50	190	240
2nd	90	150	240

(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.

Below 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 S on its first trial and an F 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 S on its first and last trials?

(g) Which table is for the study in which the subject performed better (S's are good) after an F than after an S?

Table 4

	Current		
Prev.	S	F	Total
S	50	49	99
F	50	80	130
Total	100	129	229

Table 5

	Current		
Prev.	S	F	Total
S	79	61	140
F	60	29	89
Total	139	90	229

Table 6

	Current		
Prev.	S	F	Total
S	27	63	90
F	63	147	210
Total	90	210	300

27. On each of five days next week (Monday thru Friday), Alice will shoot four free throws. Assume that Alice's shots satisfy the assumptions of Bernoulli trials with $p = 0.75$.

    (a) Compute the probability that Alice obtains a total of (exactly) three successes on any particular day. For future reference, if Alice obtains exactly three successes on a particular day, then we say that the event "Brad" has occurred.

    (b) Compute the probability that Brad occurs exactly twice next week. For future reference, if Brad occurs exactly twice during a week, we say that the event "Mel" has occurred.

    (c) Compute the probability that in the next four weeks, Mel occurs, then does not occur twice, then occurs, in that order.

28. On each of four days (Monday thru Thursday) every week for the next four weeks, Alex will shoot five free throws. Assume that Alex's shots satisfy the assumptions of Bernoulli trials with $p = 0.45$.

    (a) Compute the probability that Alex obtains a total of (exactly) two successes on any particular day. For future reference, if Alex obtains exactly two successes on a particular day, then we say that the event "Brad" has occurred.

    (b) Compute the probability that Brad occurs exactly once next week. For future reference, if Brad occurs exactly once during a week, we say that the event "Mel" has occurred.

    (c) Compute the probability that in the next four weeks, Mel occurs, then does not occur twice, then occurs, in that order.

29. On each of four days next week (Monday thru Thursday), Carl will shoot five free throws. Assume that Carl's shots satisfy the assumptions of Bernoulli trials with $p = 0.62$.

    (a) Define the event B as follows:

        On any given day next week, Carl obtains exactly three successes.

        Compute $P(B)$.

    (b) Refer to part (a). Define C to be the event:

        Next week, event B occurs on Monday and Wednesday and does not occur on Tuesday and Thursday.

        Compute $P(C)$.

    (c) Refer back to the beginning of this question (before part (a)). Define the event D as follows:

        On next Monday, Carl makes his first free throw and a total of exactly four free throws.

        Compute $P(D)$.

Section 6.3

30. Don and Eric select independent random samples from the same population. Their sample sizes are different. Don uses his data to compute a b% confidence interval for p. Eric uses his data to compute c% and d% confidence intervals for p. The exact values of $b, c,$ and

d are unimportant, but you need to know that $b < c < d$.

As a result of their computations, Don and Eric have three lower bounds and three upper bounds. Unfortunately, these six numbers became mixed up; their sorted values are below.

$$0.63 \quad 0.65 \quad 0.69 \quad 0.70 \quad 0.71 \quad 0.72$$

Given that 0.65 is one of the lower bounds and 0.69 is one of the upper bounds, reconstruct the three confidence intervals and match each interval with its confidence level.

31. Don selects a random sample from a population and Eric selects a random sample from a different population. Their sample sizes are different too. Don uses his data to compute a confidence interval for p. Eric uses his data to compute two confidence intervals for p. Their three confidence levels are, after sorting: 80%, 95%, and 99%. The confidence intervals are:

Label:	A	B	C
CI:	[0.20,0.32]	[0.21,0.35]	[0.23,0.29]

Match each confidence interval to its level and researcher; remember to match two intervals to Eric and to use all three confidence levels.

32. Mike selects a random sample from a population and uses his data to construct two confidence intervals for p. Nancy selects a random sample from the same population and uses her data to construct one confidence interval for p. The three confidence intervals are:

A	B	C
$[0.295, 0.385]$	$[0.305, 0.375]$	$[0.317, 0.383]$

The three confidence levels are 80%, 90%, and 95%.

Match each interval to its confidence level and to its researcher.

Section 6.5

33. I cast my round-corned die 1,000 times and obtained the number six 240 times. I plan to cast the die an additional 500 times. Assume that successive casts are Bernoulli trials.

    (a) Compute the point prediction of the number of times the die yields a six in the additional 500 casts.

    (b) Calculate the 90% prediction interval for the number of times die yields a six in the additional 500 casts.

    (c) I perform the additional 500 casts and obtain the number six a total of 131 times. Comment on your answers to parts (a) and (b) above.

34. Bert enjoys the dart game Cricket. To prepare for Cricket, he practices by aiming at the region of the dart board marked '20.' (For the purpose of this exercise, we won't distinguish between single, double and triple twenties.) In 220 throws, Bert hits a '20' a total of 77 times. Assume that successive throws are Bernoulli trials.

    (a) Compute the point prediction of the number of times Bert hits a '20' in his next 140 throws.

    (b) Calculate the 90% prediction interval for the number of times Bert hits a '20' in his next 140 throws.

    (c) Bert performs the additional 140 throws and hits a '20' a total of 62 times. Comment on your answers to parts (a) and (b) above.

Chapter 7

35. I select a random sample of size 180 from population 1 and obtain 63 successes. I select an independent random sample of size 300 from population 2 and obtain 87 successes.

    (a) Compute the point estimate of p_1, the proportion of successes in the first population.

(b) Compute the 95% confidence interval estimate of p_1.

(c) Compute the point estimate of $p_1 - p_2$.

(d) Compute the 90% confidence interval estimate of $p_1 - p_2$.

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

Group	S	F	Total
1	60	40	100
2	55	45	100
Total	115	85	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.

Subgroup A				Subgroup B			
Grp.	S	F	Tot.	Grp.	S	F	Tot.
1	49	21	70	1	11	19	30
2			40	2			60
Tot.			110	Tot.			90

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

Group	S	F	Total
1	42	58	100
2	48	52	100
Total	90	110	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.

Subgroup A				Subgroup B			
Grp.	S	F	Tot.	Grp.	S	F	Tot.
1	21	49	70	1	21	9	30
2			30	2			70
Total			100	Total			100

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

Group	S	F	Total
1	45	55	100
2	36	64	100
Total	81	119	200

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

Subgp A				Subgp B			
Grp.	S	F	Tot	Grp.	S	F	Tot
1	33	27	60	1	12	28	40
2			30	2			70
Tot			90	Tot			110

39. In a "Chapter 7" problem you are given the following information.

$$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 0.04 \text{ and } \sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}} = 0.06$$

One of the following is the half width of the 95% confidence interval for $p_1 - p_2$; which one is it?

(a) $1.96(0.04 + 0.06)$

(b) $1.96\sqrt{0.04 + 0.06}$

(c) $1.96[(0.04)^2 + (0.06)^2]$

(d) $1.96\sqrt{(0.04)^2 + (0.06)^2}$

Section 12.3

40. I have drawn a histogram for 200 observations. One of the rectangles has endpoints of 0.50 and 0.60, and a height of 4. For each of the three situations below, determine *how many* observations are in this class interval [0.50 to 0.60), with the usual endpoint convention.

(a) If it is a frequency histogram.

(b) If it is a relative frequency histogram.

(c) If it is a density scale histogram.

Section 12.4

41. Below are 50 sorted observations.

0.05	0.07	0.08	0.09	0.13
0.22	0.33	0.42	0.43	0.46
0.49	0.71	0.93	0.99	1.36
1.50	1.58	1.59	2.01	2.50
2.64	2.67	2.88	2.91	3.01
3.18	3.34	3.53	3.57	3.71
3.76	3.96	4.00	4.04	4.28
4.35	4.41	4.45	4.56	4.57
4.63	4.65	4.75	4.78	4.83
4.85	4.92	4.93	4.94	4.96

(a) Draw a density scale histogram of these data. Use 0.00–0.50, 0.50–2.50, 2.50–4.50 and 4.50–5.00 as your class intervals. Clearly label the height and endpoints of each of the four rectangles.

(b) Obtain the interquartile range of these data.

(c) Given the mean and standard deviation of these data are 2.76 and 1.78, respectively, determine the *proportion* of observations that are within one standard deviation of the mean. How does your proportion compare to value predicted by the empirical rule? Are you are surprised by the agreement/disagreement? Comment.

42. Data are collected from 100 subjects. The response is a measurement rounded to the nearest hundredth. For example, The smallest response value is 0.07 and the largest is 5.87. The data yield the following frequency table. Note that none of the values fall on a boundary of a class interval (i.e. none of the measurements equals 1.00, 2.00, 3.00, 4.00, or 5.00).

Class Interval	Relative Frequency
0.00–1.00	0.30
1.00–2.00	0.13
2.00–3.00	0.07
3.00–4.00	0.08
4.00–5.00	0.15
5.00–6.00	0.27

(a) Suppose that I draw a density scale histogram for these data and that one of my class intervals is 1.00–3.00. What is the height of the rectangle over this interval?

(b) The IQR is ____________

i. larger than 4

ii. smaller than 4

iii. impossible to tell whether it is larger or smaller than 4

(c) Given the mean and standard deviation of these data are 3.00 and 2.00, respectively, determine the *proportion* of observations that are within one standard deviation of the mean.

(d) Refer to part (c). Compare your answer with the empirical rule prediction and explain why your answer is not unexpected.

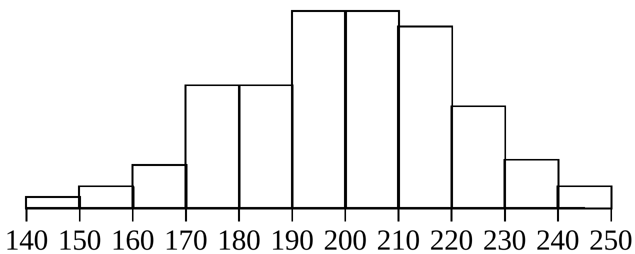
43. Data are collected from 100 subjects. The response is a measurement rounded to the nearest hundredth. For example, The smallest response value is 0.07 and the largest is 5.87. The data yield the following frequency table. Note that none of the values fall on a boundary of a class interval (i.e. none of the measurements equals 1.00, 2.00, 3.00, 4.00, or 5.00).

Class Interval	Relative Frequency
0.00–1.00	0.27
1.00–2.00	0.13
2.00–3.00	0.10
3.00–4.00	0.08
4.00–5.00	0.13
5.00–6.00	0.29

(a) Is the first quartile (Q_1): smaller than 1.00; equal to 1.00; larger than 1.00; or impossible to tell?

(b) Is the median ($\tilde{x}$): smaller than 3.00; equal to 3.00; larger than 3.00; or impossible to tell?

44. A sample of size $n = 200$ yields the histogram printed below.



The mean of these data is 200.00. The standard deviation is one of the following four numbers.

5.00 10.00 20.00 40.00

Circle the standard deviation. Explain your answer.

45. A sample of size 33 yields the following sorted data. Note that I have x-ed out $x_{(32)}$ (the second largest number). This fact will NOT prevent you from answering the questions below.

2.86	5.72	6.22	6.33	6.37	6.48
6.56	7.25	7.57	8.27	8.77	9.74
11.19	11.20	11.97	12.60	12.64	13.59
14.47	17.39	18.59	18.77	18.98	19.07
21.50	23.36	25.71	28.83	30.91	37.47
44.18	x	56.85			

Hint: The mean of these data is (exactly) 17.56 and the standard deviation is 13.23.

(a) Suppose you draw a density scale histogram of these data. One of the class intervals is 5.00–10.00; how tall is its rectangle?

(b) What proportion of these data lies within one standard deviation of the mean?

(c) Calculate the median and IQR of these data.

(d) We discover that the observation 5.72 above is really 15.72. After making this change to the data set, calculate the first quartile, median and mean.

Section 15.3

46. A random sample of size 16 is selected from a pdf. The sorted data are below.

17.9	22.0	22.1	24.2	24.9	25.6
29.1	30.6	36.1	37.5	38.7	40.3
42.3	46.0	49.0	56.1		

(Hint: $\bar{x} = 33.90$ and $s = 11.06$.)

(a) Construct a confidence interval for the median of the population. Select the confidence level and remember to report it with your answer.

(b) Construct Gosset's 90% confidence interval for the mean of the population.

(c) Test the null hypothesis that $\mu = 27$ versus the alternative that $\mu > 27$.

47. A random sample of size 25 is selected from a pdf. The sorted data are below.

36.1	46.6	51.2	52.3	52.8
56.6	57.7	58.6	59.1	59.8
60.5	66.5	66.7	66.8	66.9
68.2	70.3	76.0	76.4	77.6
79.3	81.3	89.5	101.5	102.4

(Hint: $\bar{x} = 67.23$ and $s = 15.89$.)

(a) Construct the 95% confidence interval for the median of the population.

(b) Construct Gosset's 95% confidence interval for the mean of the population.

(c) Test the null hypothesis that $\mu = 75$ versus the alternative that $\mu < 75$.

48. Each of 18 researchers selects a random sample from a given population and calculates a confidence interval for the mean. The 18 lower bounds are below.

59.0	60.9	63.4	70.7	71.1	72.1
73.3	76.2	87.9	88.2	89.7	94.0
99.1	101.6	103.7	104.2	118.7	121.9

Given that exactly four of these 18 intervals are "too large" (i.e. all numbers in the interval are larger than μ), what can you say about the value of μ?

49. Each of 18 researchers selects a random sample from a given population and calculates a confidence interval for the mean. The 18 upper bounds are below.

39.0	39.8	43.3	45.4	55.2	57.9
59.0	60.9	63.4	70.7	71.1	72.1
73.3	76.2	87.9	88.2	89.7	94.0

Given that exactly three of these 18 intervals are "too small" (i.e. all numbers in the interval are smaller than μ), what can you say about the value of μ?

50. Alma, Betty, and Chloe each calculate an 80% confidence for the mean of a population. Their intervals are Alma: $[29, 36]$; Betty: $[32, 38]$; and Chloe: $[33, 42]$. (Hint: For each of the questions below, the correct answer is not any of the above three intervals.)

    (a) Suppose Nature announces, "All three intervals are correct." Given this information, what is the narrowest interval known to contain μ?

    (b) Suppose Nature announces, "The intervals of Alma and Betty are correct, but Chloe's is incorrect." Given this information, what is the narrowest interval known to contain μ?

    (c) Suppose Nature announces, "The intervals of Alma and Chloe are correct, but Betty's is incorrect." Given this information, what is the narrowest interval known to contain μ?

51. 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.

$$[9, 13], [11, 15], [12, 17], \text{ and } [6, 10].$$

    (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.)

52. 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.

$$[5, 20], [8, 17], [10, 23], \text{ and } [15, 30].$$

    (a) Nature announces, "All four intervals are correct." Given this information, what is the smallest interval that is known to contain the median?

    (b) Nature announces, "Two of the intervals are correct, and two intervals are too small." Given this information, what is the smallest interval that is known to contain the median?

(c) Nature announces, "Two of the intervals are correct, and two intervals are too large." Given this information, what is the smallest interval that is known to contain the median?

(d) Nature announces, "Two of the intervals are correct, one interval is too small, and one interval is too large." Explain why we can conclude that Nature is mistaken. Be precise.

53. Fifteen researchers select random samples from the same population. Each researcher computes a confidence interval for the mean of the population. Thus, each researcher is estimating the same number. Unfortunately, the lower and upper bounds of the 15 confidence intervals became disconnected. The 15 lower bounds, sorted, are below.

14.39	14.96	15.50	16.43	16.51
17.19	17.76	18.51	18.56	18.94
19.08	20.51	22.05	22.09	22.55

The 15 upper bounds, sorted, are below.

17.40	20.64	21.12	21.48	21.50
21.60	21.86	22.33	22.40	22.49
22.86	23.00	23.49	24.34	26.48

Note that the information given below in part (a) **must not be used** in parts (b), (c) or (d).

(a) Given that $\mu = 20$, how many of the confidence intervals are correct?

(b) Given that exactly three of the confidence intervals are too large, what can you conclude about the value of μ?

(c) Refer to part (b). Suppose we are now told that exactly one of the confidence intervals is too small. Adding this new information to your conclusion in part (b), what can you now conclude about the value of μ?

(d) Now take μ to be unknown. What is the largest number of intervals that could be correct? What values of μ give this largest number?

Section 16.2

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

52.6	55.1	55.2	55.4	55.5	55.8
58.6	59.0	59.6	60.2	66.8	67.8
69.1	70.3	72.1	74.5		

Hint: The mean and standard deviation of these numbers are 61.73 and 7.16.

Below are the sorted data from the second population.

44.1	48.8	59.3	60.3	66.6	69.0

Hint: The mean and standard deviation of these numbers are 58.02 and 9.80.

(a) Calculate the 95% confidence interval for the difference of the means of the two populations.

(b) Calculate Gosset's 90% confidence interval for the mean of the first population.

(c) For the first population, find the P-value for testing the null hypothesis that $\mu = 58$ versus the alternative that $\mu > 58$. Show your work.

(d) Calculate a confidence interval for the median of the first population. Select your confidence level and report it with your answer.

(e) Suppose now that the two samples came from the same population. Thus, we will combine the two samples into one random sample. Use this combined sample to obtain the 95% confidence interval for the median of the population.

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

Pop.	Mean	Stand. Dev.	Sample size
1	38.00	8.00	15
2	32.00	6.00	9

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

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

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

Pop.	Mean	Stand. Dev.	Sample size
1	28.00	6.00	6
2	22.00	4.00	12

(a) Construct the 90% confidence interval for $\mu_X - \mu_Y$.

(b) Obtain the P-value for the alternative $\mu_X \neq \mu_Y$. Make sure to specify which reference curve you use.

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

- $\bar{x} = 32.00, s_X = 6.00$, and $n_1 = 15$
- $\bar{y} = 28.00, s_Y^2 = 29.00$, and $n_2 = 7$

Calculate s_p.

58. 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.

(a) The alternative is $\mu_X > \mu_Y$; the value of the test statistic is $t_1 = 2.080$; the sample sizes are 13 and 9.

(b) The alternative is $\mu_X < \mu_Y$; the value of the test statistic is $t_1 = -3.012$; the sample sizes are 8 and 6.

(c) The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is $t_1 = 2.950$; the sample sizes are 11 and 8.

Section 8.2

59. 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.)

	B	B^c	Total
A	190	10	200
A^c	50	750	800
Total	240	760	1000

(a) What proportion of the population has the disease?

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

(c) Of those who have the disease, what proportion would test negative?

(d) Of those who would test positive, what proportion is free of the disease?

(e) What proportion of the population has the disease or would test positive?

60. 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.)

	B	B^c	Total
A	49	7	56
A^c	14	210	224
Total	63	217	280

(a) What proportion of the population has the disease?

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

(c) Of those who have the disease, what proportion would test negative?

(d) Of those who would test negative, what proportion is free of the disease?

(e) What proportion of the population has the disease or would test positive?

61. Casey (my dog) is looking out the window. There is a 20% chance that she will see one or more squirrels in the next 10 minutes. Given that she sees one or more squirrels, there is an 90% chance that Casey will bark during the time period. Given that she sees no squirrels, there is a 35% chance that Casey will bark during the time period.

(a) What is the probability that Casey will bark during the next 10 minutes?

(b) Given that Casey barks during the next 10 minutes, what is the probability that she saw one or more squirrels?

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

Eighty percent of the defendants are, in fact, guilty. Given that the defendant is guilty, there is an 87.5 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.

(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?

63. I love to watch movies. And in every movie I watch, I select one (male) actor to be **the hero** (for example, the bald guy in *The Princess Bride*; Buzz Lightyear in *Toy Story*). My favorite actor by far is Bruce Willis (BW) (note his initials; a coincidence? I don't think so!). As a result, if BW is in a movie, he is *always* the hero.

I have noticed, however, a disturbing trend in recent years. It seems that, as BW ages, directors/screenwriters like to have him die! Sometimes, in quite horrible ways! Thus, I collected the data described below.

I made a list of my 100 favorite movies. For each movie, I identified one of the (male) actors to be the hero. Then I determined whether the hero died in the movie. Here is what I found.

- BW appears (and is, hence, the hero) in 20 of the movies.
- The hero dies in 40% of the movies in which BW appears.
- The hero dies in 5% of the movies in which BW does not appear.

Consider the chance mechanism of selecting one movie at random from the list of my 100 favorite movies.

(a) What is the probability I select a movie in which the hero dies?

(b) Given that the hero dies, what is the probability he is BW?

(c) Given that the hero does not die, what is the probability he is BW?

Section 13.3

64. Eighty students take midterm and final exams. The scores on the midterm have a mean of 45 and a standard deviation of 6. The scores on the final have a mean of 60 and a standard deviation of 9. The correlation coefficient of the set of two scores is 0.60.

(a) Obtain the regression line for using the midterm score to predict the final score.

(b) Obtain the regression line for using the final score to predict the midterm score.

Section 13.4

65. Fran calculates the regression line for her data and obtains $\hat{y} = 15 + 2.5x$.

(a) One of the subjects, Mona, has $x = 50$ and $y = 147$. Calculate the predicted value for Mona.

(b) Refer to part (a). Calculate the residual for Mona.

(c) One of the subjects, Ralph, has $y = 160$, which is 15 units smaller than his predicted response. What is Ralph's x?

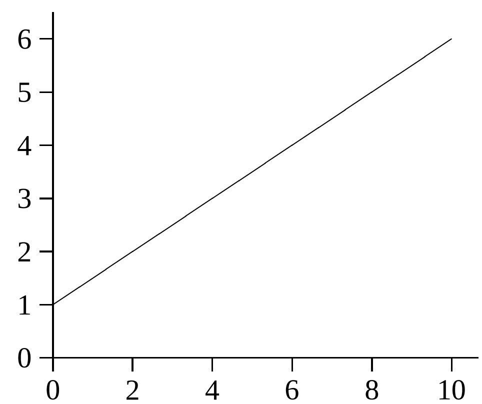
66. Sixty students take two midterm exams. The scores on the first midterm have a mean of 55.00 and a standard deviation of 10.00. The scores on the second midterm have a mean of 63.00 and a standard deviation of 15.00. The correlation coefficient of the set of two scores is 0.62. One of these students, Barbara, scored 45 on the first midterm and 65 on the second midterm.

   (a) Obtain the regression line for using the first midterm score to predict the second midterm score.

   (b) Use your answer to part (a) to obtain a predicted score for Barbara.

   (c) Calculate Barbara's residual.

67. Below is a coordinate system with the regression line $\hat{y} = 1 + 0.5x$.



   (a) Locate the point that has $x = 8$ and $y = 2$; put an A at that point.

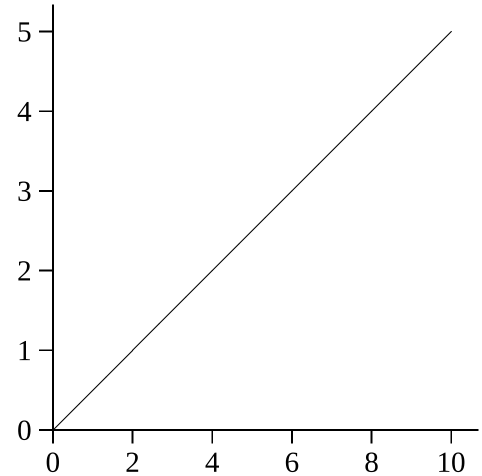
   (b) Locate the point that has $x = 4$ and $e = -2$; put a B at that point.

   (c) Locate the point that has $y = 6$ and $e = 3$; put a C at that point.

   (d) Draw the line that represents all points for which $e = +1$.

   (e) Given that $\bar{x} = 5$, what is the value of $\bar{y}$?

68. Below is a coordinate system with the regression line $\hat{y} = 0.5x$.



   (a) Locate the point that has $x = 4$ and $y = 3$; 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 $y = 2$ and $e = -3$; put a C at that point.

   (d) Draw the line that represents all points for which $e = +1$.

   (e) Given that $\bar{x} = 6$, what is the value of $\bar{y}$?

69. A regression analysis yields $\bar{x} = 40$ and $\bar{y} = 80$. In addition, one of the subjects, Sally, has $x = 50$, $y = 65$ and $e = -5$. Determine the equation of the regression line.

70. A regression analysis yields $\bar{x} = 40$ and $\bar{y} = 90$. In addition, one of the subjects, Sam, scores five above the mean on both variables. Also, she scores eight points higher than predicted by the regression line.

   Determine the equation of the regression line.