Extra Exercises; Statistics 301;
Professor Wardrop

Section 5.2

1. Four cards are selected at random with replacement from a dichotomous box with $p = 0.46$. Calculate the probability of each of the following sequence of outcomes.
   (a) 1100; i.e. successes on the first two trials and failures on the last two trials.
   (b) 1001.
   (c) 1110.
   (d) 0010.

2. Five cards are selected at random with replacement from a dichotomous box with $p = 0.70$. Calculate the probability of each of the following sequence of outcomes.
   (a) 11011.
   (b) 10111.
   (c) 11100.
   (d) 00110.

3. Refer to the chance mechanism described in exercise 1. Let X denote the total number of successes obtained. Calculate the following probabilities.
   (a) $P(X = 3)$.
   (b) $P(X = 2)$.

4. Refer to the chance mechanism described in exercise 2. Let X denote the total number of successes obtained. Calculate the following probabilities.
   (a) $P(X = 3)$.
   (b) $P(X = 4)$.

5. Refer to X defined in exercise 3. Calculate the mean and standard deviation of X.

6. Refer to X defined in exercise 4. Calculate the mean and standard deviation of X.

Section 5.3

7. 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.

8. 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.

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

   A This table provides evidence that p increased over the course of the study.

   B This table provides evidence that p decreased over the course of the study.

   C This table provides no evidence that p changed over the course of the study.

Table 1

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

Table 2

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

Table 3

Half	S	F	Total
1st	130	110	240
2nd	110	130	240

(b) 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 one of the following four questions.

Which table is for the study that had an F on its first trial and an S on its last trial?

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

Which table is for the study that had an F on its first and last trials?

Which table is for the study in which the subject performed the same after an S as after an F?

Table 4

	Current		
Prev.	S	F	Total
S	53	52	105
F	53	77	130
Total	106	129	235

Table 5

	Current		
Prev.	S	F	Total
S	74	70	144
F	69	31	100
Total	143	101	244

Table 6

	Current		
Prev.	S	F	Total
S	9	36	45
F	36	144	180
Total	45	180	225

10. 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

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

Table 5

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

Table 6

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

11. 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.

12. 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.

13. 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

14. The purpose of this question is to connect the material on confidence intervals in Chapter 6 to the material on computer simulation in Chapter 3.

    Recall the Crohn's study. It can be shown that there are

    $$2.093 \times 10^{20}$$

    possible assignments. (Because I don't recall the proper words for numbers bigger than trillions, my best verbal description of this number is 209.3 million trillions; it's big!) Consider a population box with N equal to this number; i.e. one card for each possible assignment. Now, of course, to use Chapter 6 we need a '1' or a '0' on each card. Here is how we proceed.

    On the assumption the Skeptic is correct, each assignment will yield a number x for the observed value of the test statistic. Recall that the actual assignment gave $x = 0.27$ and remember that the P-value was $P(X \geq 0.27)$. In the population box mark a card a '1' if it's x is 0.27 or larger; otherwise, mark it a '0.' Next, note that the p for the population box is the proportion of assignments that would give an $x \geq 0.27$; that is, the p for the population box equals the P-value for the hypothesis test. Thus, estimating p is the same as estimating (or as we called it in Chapter 3, approximating) the P-value.

    The simulation experiment is simply a sample of $n = 10,000$ cards from the population box. From page 1, column 2 of the Course Notes, we know that $\hat{p} = 0.0193$.

    Calculate the 99% confidence interval for the P-value. Is it correct?

15. Refer to the previous question. Use the information on page 1, column 2 of the Course Notes to calculate the 98% confidence interval for the P-value for the Ballerina study. Is your interval correct? (Note: Although it does not enter into your solution, for completeness let me remark that $N = 126$ trillion.)

16. Abby and Beth select independent random samples from the same population. Their sample sizes are different. Beth uses her data to compute a 90% confidence interval for p. Abby uses her data to compute 80% and 95% confidence intervals for p.

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

    $$0.345 \quad 0.363 \quad 0.364 \quad 0.436 \quad 0.455 \quad 0.477$$

    Given that 0.363 is one of the lower bounds and 0.455 is one of the upper bounds, reconstruct the three confidence intervals and match each interval with its confidence level.

17. 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.

18. Carla selects a random sample from a population and Deb selects a random sample from a different population. Their sample sizes are different too. Carla uses her data to compute a confidence interval for p. Deb uses her 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:	[0.185,0.295]	[0.182,0.318]	[0.170,0.310]

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

19. 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.

20. 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.

21. Norma selects a random sample from a population and uses her data to construct three confidence intervals for p. The three confidence intervals are:

A	B	C
$[0.295, 0.385]$	$[0.305, 0.375]$	$[0.317, 0.363]$

The three confidence levels are selected from our usual five: 80%, 90%, 95%, 98% and 99%.

Match each interval to its confidence level.

Section 6.5

22. 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.

23. 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

24. 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.

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

25. I select a random sample of size 500 from population 1 and obtain 200 successes. I select an independent random sample of size 600 from population 2 and obtain 270 successes.

Compute the 80% confidence interval estimate of $p_1 - p_2$.

26. 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

27. 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

28. An observational study yields the following “collapsed table.”

Group	S	F	Total
1	100	150	250
2	95	155	250
Total	195	305	500

Below are two component tables for these data.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	54	96	150	1	46	54	100
2			100	2			150
Tot			250	Tot			250

Complete these tables so that Simpson’s Paradox is occurring **or** explain why Simpson’s Paradox *cannot* occur for these data. You must present computations to justify your answer.

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

Group	S	F	Total
1	70	130	200
2	38	62	100
Total	108	192	300

Below are two component tables for these data.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	45	105	150	1	25	25	50
2			40	2			60
Tot			190	Tot			110

Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. You must present computations to justify your answer.

30. 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) $1.96\sqrt{(0.04)^2 + (0.06)^2}$