

**Extra Exercises; Statistics 301;  
Professor Wardrop  
Spring 2008**

**Section 12.3**

1. 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.

2. I have drawn a histogram for 400 observations. One of the rectangles has endpoints of 0.50 and 0.60, and a height of 0.25. 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.

3. 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) Calculate the median of these data.
- (b) Draw a frequency histogram of these data. Use 0.00–1.00, 1.00–2.00, and so on, as your class intervals. Clearly label the height and endpoints of each of the five rectangles.

(c) 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.

4. Below are 100 sorted observations.

0.01	0.02	0.02	0.05	0.06
0.09	0.14	0.14	0.15	0.15
0.15	0.16	0.20	0.21	0.24
0.26	0.31	0.32	0.34	0.38
0.39	0.40	0.47	0.47	0.49
0.49	0.51	0.53	0.53	0.55
0.59	0.61	0.62	0.63	0.64
0.78	0.82	0.86	0.93	0.94
0.97	0.99	1.03	1.04	1.05
1.18	1.19	1.29	1.35	1.36
1.37	1.39	1.46	1.48	1.51
1.52	1.59	1.61	1.70	1.70
1.75	1.79	1.79	1.79	1.81
1.86	1.97	2.09	2.11	2.13
2.16	2.23	2.38	2.39	2.49
2.52	2.56	2.71	2.88	3.00
3.13	3.37	3.49	3.56	3.58
3.61	3.85	4.05	4.20	4.22
4.40	4.52	4.74	5.16	5.27
5.37	5.48	5.90	6.75	6.82

- (a) Calculate the median of these data.
- (b) Draw a relative frequency histogram of these data. Use five intervals of equal width, beginning at 0.00 and ending at 7.50. Clearly label the height and endpoints of each of the five rectangles.
- (c) Draw a density scale histogram of these data. Use 0.00–0.50, 0.50–1.00, 1.00–2.00, 2.00–4.00 and 4.00–7.00 as your class intervals. Clearly label the height and endpoints of each of the four rectangles.

## Section 12.4

5. Refer to the data in exercise 3 above.
- Calculate the first and third quartiles of these data.
  - 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 surprised by the agreement/disagreement? Comment.
6. Refer to the data in exercise 4 above.
- Calculate the first and third quartiles of these data.
  - Given the mean and standard deviation of these data are 1.78 and 1.65, 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 surprised by the agreement/disagreement? Comment.
7. 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.

- What proportion of these data lies in the interval  $[\bar{x} - s, \bar{x} + 2s]$ ?
- Calculate the median and quartiles of these data.

- Suppose 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.

8. A sample of size  $n = 28$  yields the following sorted data. Note that the largest number in the list has been replaced by the letter 'y.'

384	420	422	456	462	466	486
492	494	518	520	572	576	580
594	618	622	630	642	644	646
650	686	712	754	764	790	y

**Hint:** The mean of these data is 586.0.

- Calculate the median and the first quartile of these data.
- Suppose we discover that the observation 384 is an error. Recalculate the median, the first quartile and the mean after deleting the observation 384.

## Section 15.3

9. 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  $s = 11.06$ .)

- Construct a confidence interval for the median of the population. Select the confidence level and remember to report it with your answer.
- Construct Gosset's 90% confidence interval for the mean of the population.
- Test the null hypothesis that  $\mu = 27$  versus the alternative that  $\mu > 27$ .

10. 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  $s = 15.89$ .)

- Construct the 95% confidence interval for the median of the population.
  - Construct Gosset's 95% confidence interval for the mean of the population.
  - Test the null hypothesis that  $\mu = 75$  versus the alternative that  $\mu < 75$ .
11. 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  $\mu$ ), what can you say about the value of  $\mu$ ?
- Given that at least three of these 18 intervals are "too large" what can you say about the value of  $\mu$ ?

12. 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  $\mu$ ), what can you say about the value of  $\mu$ ?
- Given that at least five of these 18 intervals are "too small" what can you say about the value of  $\mu$ ?

13. 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.)

- Suppose Nature announces, "All three intervals are correct." Given this information, what is the narrowest interval known to contain  $\mu$ ?
- 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  $\mu$ ?
- 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  $\mu$ ?

14. 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], and [6, 10].

- 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.)
- 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.)

15. 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], 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.
16. 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  $\mu$ ?
- (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  $\mu$ ?
- (d) Now take  $\mu$  to be unknown. What is the largest number of intervals that could be correct? What values of  $\mu$  give this largest number?

### Section 16.2

17. 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.

18. 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  $\mu_X \neq \mu_Y$ .
19. 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.

20. 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  $s_p$ .

21. 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.

22. 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  $t_1 = 1.725$ ; the sample sizes are 15 and 7.
- (b) The alternative is  $\mu_X < \mu_Y$ ; the value of the test statistic is  $t_1 = -1.012$ ; the sample sizes are 18 and 6.
- (c) The alternative is  $\mu_X \neq \mu_Y$ ; the value of the test statistic is  $t_1 = -2.634$ ; the sample sizes are 9 and 8.

### Section 8.2

23. 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?
24. 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?
25. 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?
26. 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?
27. 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.

- What is the probability I select a movie in which the hero dies?
- Given that the hero dies, what is the probability he is BW?
- Given that the hero does not die, what is the probability he is BW?

### Section 13.3

28. 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.
- Obtain the regression line for using the midterm score to predict the final score.
  - Obtain the regression line for using the final score to predict the midterm score.

### Section 13.4

29. Fran calculates the regression line for her data and obtains  $\hat{y} = 15 + 2.5x$ .
- One of the subjects, Mona, has  $x = 50$  and  $y = 147$ . Calculate the predicted value for Mona.
  - Refer to part (a). Calculate the residual for Mona.
  - One of the subjects, Ralph, has  $y = 160$ , which is 15 units smaller than his predicted response. What is Ralph's  $x$ ?

30. 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.

- Obtain the regression line for using the first midterm score to predict the second midterm score.
- Use your answer to part (a) to obtain a predicted score for Barbara.
- Calculate Barbara's residual.

31. Recall that for a regression analysis,

$$\sum e = 0 \text{ and } \sum xe = 0.$$

Use these facts and the chart below to determine the numerical values of  $e_1$  and  $e_3$ .

$x$	$e$
0	5
1	$e_1$
2	-3
3	$e_3$

32. A regression analysis yields:

$$s_Y = 10.0, s_X = 4.0 \text{ and } b_1 = 3.0.$$

Explain why these values are not possible.

33. 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.
34. 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.