Practice Exam Questions and Solutions
for Chapters 6, 7, 12, 15, 16, 8 and 13
Statistics 301, Professor Wardrop

Chapter 6

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

2. A random sample of size $n = 600$ yields 228 successes. Calculate the 98% confidence interval for p.

3. 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).

4. Sally enjoys playing Diet-Rite-Pong. Last week she performed 100 trials and obtained 32 successes. Next week she plans to perform 350 trials. Assume that Sally's trials satisfy the assumptions of Bernoulli trials.

   (a) Calculate the point prediction of the number of successes that Sally will obtain next week.

   (b) Calculate the 80% prediction interval for the number of successes Sally will obtain next week.

   (c) It turns out that next week Sally obtains 125 successes. Given this information, comment on your answers in parts (a) and (b).

5. 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.

6. Enid selects one random sample from a population and calculates three confidence intervals for p. Her intervals are below.

A	B	C
$\hat{p} \pm 0.0201$	$\hat{p} \pm 0.0308$	$\hat{p} \pm 0.0365$

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

Chapter 7

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

Group	S	F	Total
1	72	228	300
2	88	212	300
Total	160	440	600

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	30	30	60	1	42	198	240
2			120	2			180
Tot			180	Tot			420

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

Group	S	F	Total
1	100	100	200
2	156	244	400
Total	256	344	600

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	70	30	100	1	30	70	100
2			80	2			320
Tot			180	Tot			420

9. An observational study yields the following collapsed table.

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

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	5	15	20	1	36	44	80
2			40	2			60
Tot			60	Tot			140

10. An observational study yields the following collapsed table.

Group	S	F	Total
1	100	150	250
2	42	58	100
Total	142	208	350

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	57	93	150	1	43	57	100
2			60	2			40
Tot			210	Tot			140

11. In a 'Chapter 7' problem, you are given the following information:

$$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 0.025, \text{ and}$$

$$\frac{\hat{p}_2\hat{q}_2}{n_2} = 0.005.$$

Calculate the half-width of the 90% confidence interval for $p_1 - p_2$.

12. In a 'Chapter 7' problem, you are given the following information:

$$1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 0.0679, \text{ and}$$

$$1.645\sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}} = 0.0409.$$

Calculate the half-width of the 80% confidence interval for $p_1 - p_2$.

Chapter 12

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

14.1	46.0	49.3	53.0	54.2	54.7	54.7
54.7	54.8	55.4	57.6	58.2	58.3	58.7
58.9	60.8	60.9	61.0	61.1	63.0	64.3
65.6	66.3	66.6	67.0	67.9	70.1	70.3
72.1	72.4	72.9	73.5	74.2	75.3	75.4
75.9	76.5	77.0	$x_{(39)}$	88.9		

(a) Construct the density scale histogram for these data using the class intervals: $[10, 40], [40, 50], [50, 60], [60, 75]$ and $[75, 90]$.

(b) Calculate the first quartile and median of these data.

(c) Calculate the 35th and 62nd percentiles of these data.

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

(e) Suppose that we learn that 14.1 is a misprint. The actual observation is 64.1. Recalculate the mean after changing the 14.1 to 64.1. Also, recalculate the first quartile and median.

14. A sample of size 50 yields the following sorted data.

0.01	0.04	0.04	0.06	0.08	0.10	0.12
0.14	0.23	0.27	0.29	0.39	0.44	0.48
0.50	0.58	0.58	0.62	0.70	0.71	0.72
0.73	0.76	0.78	0.88	1.00	1.08	1.09
1.10	1.10	1.14	1.16	1.30	1.30	1.31
1.57	1.60	1.74	1.80	1.85	1.88	1.96
2.22	2.27	2.45	2.96	3.44	3.86	4.05
4.52						

Hint: For these data, $\bar{x} = 1.20$ and $s = 1.09$.

(a) Construct the density scale histogram for these data using the class intervals: $[0.00, 0.50], [0.50, 1.00], [1.00, 2.00]$ and $[2.00, 5.00]$.

(b) Calculate the median of these data.

(c) Calculate the 42nd and 77th percentiles of these data.

(d) What proportion of these data lie within one standard deviation of the mean?

(e) Calculate Q_1, Q_3 and the mean after deleting the two largest observations.

Chapter 15

15. 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.

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

61.6	64.7	65.2	67.2	68.4	69.8
72.5	73.4	76.1	76.5	77.2	78.5
80.0	81.1	84.1	86.6	90.4	91.0

Hint: The mean and standard deviation of these numbers are 75.79 and 8.76.

Below are the data from the second population.

69.4	67.6	74.2	72.1	74.0	67.9
78.7	83.6				

Hint: The mean and standard deviation of these numbers are 73.44 and 5.54.

(a) Calculate Gosset's 98% 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 80% confidence interval for the median of the population.

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

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

$$[10, 52], [24, 69], [37, 62] \text{ and } [58, 81].$$

(a) Nature announces, "Two of the intervals are correct, one interval is too small and one interval is too large." Given this information, determine all possible values of the mean. **Do not explain.**

(b) Nature announces, "Exactly three of the intervals are correct." Given this information, determine all possible values of the mean. **Do not explain.**

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

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

$$[43, 82], [49, 92], [57, 95] \text{ and } [86, 112].$$

(a) Nature announces, "Two of the intervals are correct, one interval is too small and one interval is too large." Given this information, determine all possible values of the mean. **Do not explain.**

(b) Nature announces, "Exactly three of the intervals are correct." Given this information, determine all possible values of the mean. **Do not explain.**

19. Eighteen 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 18 confidence intervals became disconnected. The 18 lower bounds, sorted, are below.

153	163	188	198	245	246
260	284	288	307	308	322
334	344	360	375	393	409

The 18 upper bounds, sorted, are below.

294	326	331	345	351	364
372	373	381	394	396	401
403	408	411	439	445	460

If you don't remember what it means for a confidence interval to be 'too small' or 'too large,' see question 11.

(a) Given that $\mu = 300$, how many of the confidence intervals are correct?

(b) Suppose we are told that exactly two of the intervals are too large. Determine all possible values of μ.

(c) Suppose we are told that at least five of the intervals are too small. Determine all possible values of μ.

(d) It could happen that all of the confidence intervals are incorrect; for example, if $\mu = 0$ then one can conclude that all of the intervals are too large. It can also be shown that it is impossible for all 18 intervals to be correct. Here is the two part question for you:

> What is the largest number of these intervals that could possibly be correct? Which values of μ will achieve this maximum number of correct intervals?

20. Twenty-four 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 24 confidence intervals became disconnected. The 24 lower bounds, sorted, are below.

328	465	501	513	550	560
561	564	577	580	581	605
610	633	634	639	642	646
649	695	720	756	794	802

The 24 upper bounds, sorted, are below.

518	535	565	568	592	598
612	645	655	656	661	666
688	693	701	752	759	772
781	806	817	837	862	873

If you don't remember what it means for a confidence interval to be 'too small' or 'too large,' see question 12.

(a) Given that $\mu = 600$, how many of the confidence intervals are correct?

(b) Suppose we are told that exactly three of the intervals are too small. Determine all possible values of μ.

(c) Suppose we are told that at least four of the intervals are too large. Determine all possible values of μ.

(d) It could happen that all of the confidence intervals are incorrect; for example, if $\mu = 0$ then one can conclude that all of the intervals are too large. It can also be shown that it is impossible for all 24 intervals to be correct. Here is the two part question for you:

> What is the largest number of these intervals that could possibly be correct? Which values of μ will achieve this maximum number of correct intervals?

Chapter 16

21. 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) Calculate s_p.

(b) Calculate the 90% CI for $\mu_X - \mu_Y$. Use Case 1.

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

- $\bar{x} = 22.50$, $s_X = 3.75$ and $n_1 = 18$
- $\bar{y} = 16.25$, $s_Y = 8.50$ and $n_2 = 6$

(a) Calculate s_p.

(b) Calculate the 98% CI for $\mu_X - \mu_Y$. Use Case 1.

23. 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 1.840; the sample sizes are 5 and 5.

    (b) The alternative is $\mu_X < \mu_Y$; the value of the test statistic is -3.150; the sample sizes are 6 and 7.

    (c) The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is 1.341; the sample sizes are 5 and 12.

    (d) The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is -0.641; the sample sizes are 12 and 12.

24. 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 0.690; the sample sizes are 4 and 4.

    (b) The alternative is $\mu_X < \mu_Y$; the value of the test statistic is -1.796; the sample sizes are 6 and 7.

    (c) The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is 1.850; the sample sizes are 7 and 14.

    (d) The alternative is $\mu_X \neq \mu_Y$; the value of the test statistic is -3.641; the sample sizes are 4 and 12.

25. Mike performs a study with $n_1 = 10$ and $n_2 = 6$. Using Case 1 from Section 16.2, he calculates an 80% CI for $\mu_X - \mu_Y$ and obtains:

$$[6.000, 14.000].$$

Calculate the 95% CI for $\mu_X - \mu_Y$ for Mike's data.

26. Maria performs a study with $n_1 = 14$ and $n_2 = 12$. Using Case 1 from Section 16.2, she calculates a 90% CI for $\mu_X - \mu_Y$ and obtains:

$$[9.500, 18.500].$$

Calculate the 99% CI for $\mu_X - \mu_Y$ for Maria's data.

Chapter 8

27. Below is the table of population counts for a disease and its screening test. (Recall that A means the disease is present and B means the screening test is positive.) On parts (a)–(e) below, report your answers as a decimal to three digits of precision, for example 0.231.

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

    (a) What proportion of the population is free of 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) What proportion of the population would receive an incorrect screening test result?

    (e) Of those who would receive a correct screening test result, what proportion would receive a correct negative?

28. 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.) To receive full credit, you must report your answers as a decimal to three digits of precision; for example, $571/1402 = 0.407$.

	B	B^c	Total
A	422	78	500
A^c	156	2253	2409
Total	578	2331	2909

(a) What proportion of the population would test positive?

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

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

(d) Of those who do not have the disease, what proportion would test positive?

(e) What proportion of the population would receive a correct screening test result?

(f) Of those who would receive an incorrect screening test result, what proportion would receive a false negative?

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

29. (Hypothetical data.) A company has 1,500 employees. You are given the following information:

- Sixty-four percent of the employees are female.
- Eighty percent of the female employees are parents.
- Forty percent of the male employees are parents.

(a) Create the table of population counts for this 'disease' and 'screening test.' Be sure to label the rows and columns of your table.

(b) An employee is selected at random from the company. Given that the employee is a parent, calculate the probability that the employee is female.

30. (Hypothetical data.) I have seen 600 movies that have been reviewed by Roger Ebert. I enjoyed 60% of these movies and I did not enjoy the remaining 40% of these movies.

Of the movies I enjoyed, 55% were recommended by Ebert. Of the movies I did not enjoy, 90% were not recommended by Ebert.

(a) What proportion of these 600 movies were recommended by Ebert?

(b) Suppose that one of these 600 movies will be selected at random. Given that Ebert recommended the movie, what is the probability that it is a movie that I enjoyed?

(c) I will say that Ebert and I disagreed about a movie if either of the following occurs: He recommended it and I did not enjoy it; or I enjoyed it and he did not recommend it. Of these 600 movies, for how many did Ebert and I disagree?

Chapter 13

31. A regression analysis yields the line

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

One of the subjects, Racheal, has $x = 60$ and $y = 52$. Another subject, Ralph, has $\hat{y} = 56$.

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

(b) Calculate Racheal's residual.

(c) Calculate Ralph's x.

32. A regression analysis yields the line

$$\hat{y} = 52 - 1.9x.$$

One of the subjects, Connie, has $x = 10$ and $y = 42$. Another subject, Craig, has $\hat{y} = 14$.

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

(b) Calculate Connie's residual.

(c) Calculate Craig's x.

33. 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.

34. Students take two midterm exams. On the first midterm exam, the mean score is 60.0 and the standard deviation is 15.00. On the second midterm exam, the mean score is 70.0 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 midterm exam score to predict the first midterm exam score.

35. 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.

36. (Hypothetical data.) Tom measures the height and weight of a group of 900 college men.

    One man, Frank, is 66 inches tall and weighs 160 pounds.

    - Frank is two inches shorter than the group mean.
    - Frank's weight is equal to the group mean.
    - Tom uses his data to obtain the regression line for using height to predict weight. According to his line, Frank weighs 6 pounds more than predicted.

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

Solutions

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

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

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

2. First, $\hat{p} = 228/600 = 0.380$. The 98% confidence interval is

$$0.380 \pm 2.326\sqrt{\frac{0.38(0.62)}{600}} =$$

$$0.380 \pm 0.046 = [0.334, 0.426].$$

3. (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.

4. (a) First, $\hat{p} = 32/100 = 0.32$. Thus, the point prediction is

$$m\hat{p} = 350(0.32) = 112.$$

   (b) The 80% prediction interval is $112\pm$

$$1.282\sqrt{112(0.68)}\sqrt{1 + (350/100)} =$$

$$112 \pm 23.7 = [88, 136],$$

   after rounding.

   (c) With 125 successes, the point prediction is too small by 13, but the prediction interval is correct b/c 125 is between 88 and 136.

5. 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%.

6. Based on the half-widths of the CIs, A has the smallest confidence level and C has the largest. But here is the key. The CIs differ only in which z they use, namely three of the following: 1.282, 1.645, 1.96, 2.326 and 2.576.

   The ratio of the half-widths of B to A is:

$$0.0308/0.0201 = 1.5323$$

   and the ratio for C to A is:

$$0.0365/0.0201 = 1.8159.$$

   These 'match' the ratios of

$$1.96/1.282 = 1.5289$$

   and

$$2.326/1.282 = 1.8143.$$

   Thus, the levels for A, B and C are 80%, 95% and 98%, respectively.

7. In the collapsed table, $\hat{p}_1 < \hat{p}_2$. To get a reversal, we need

$$c/120 < 30/60, \text{ or } c < 60 \text{ in Sbgp A and}$$

$$c/180 < 42/240, \text{ or } c < 31.5$$

   in Subgroup B. Also, the two c's must sum to the 88 in the collapsed table. There are three possible answers: 59 and 29; 58 and 30; 57 and 31.

8. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need

$$c/80 > 70/100, \text{ or } c > 56 \text{ in Sbgp A and}$$

$$c/320 > 30/100, \text{ or } c > 96$$

   in Subgroup B. Also, the two c's must sum to the 156 in the collapsed table. Possible answers are: 57 and 99; 58 and 98; 59 and 97.

9. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need

$$c/40 > 5/20, \text{ or } c > 10 \text{ in Sbgp A and}$$

$$c/60 > 36/80, \text{ or } c > 27$$

   in Subgroup B. Also, the two c's must sum to the 35 in the collapsed table. This combination of restrictions is impossible.

10. In the collapsed table, $\hat{p}_1 < \hat{p}_2$. To get a reversal, we need

$$c/60 < 57/150, \text{ or } c < 22.8 \text{ in Sbgp A and}$$

$$c/40 < 43/100, \text{ or } c < 17.2$$

    in Subgroup B. Also, the two c's must sum to the 42 in the collapsed table. This combination of restrictions is impossible.

11. The half-width is

$$1.645\sqrt{(0.025)^2 + 0.005} = 0.123.$$

12. First,

$$\frac{\hat{p}_1\hat{q}_1}{n_1} = (\frac{0.0679}{1.96})^2 = 0.0012, \text{ and}$$

$$\frac{\hat{p}_2\hat{q}_2}{n_2} = (\frac{0.0409}{1.645})^2 = 0.0006.$$

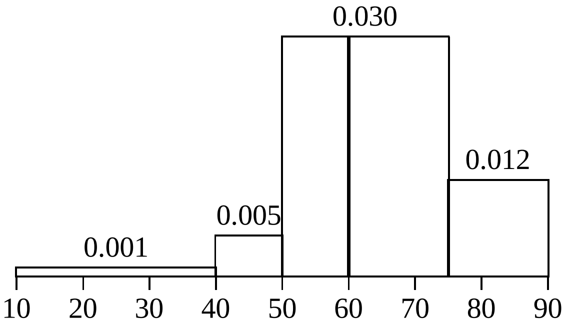
    Thus, the half-width is

$$1.282\sqrt{0.0012 + 0.0006} = 0.0544.$$

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

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

The density scale histogram is below.



(b) The median is $(63.0 + 64.3)/2 = 63.65$. The first quartile is $(55.4 + 57.6)/2 = 56.5$.

(c) For the 35th percentile: $0.35(40) = 14$ is an integer. Thus, the 35th percentile is the mean of the numbers in positions 14 and 15:

$$(58.7 + 58.9)/2 = 58.8.$$

For the 62nd percentile: $0.62(40) = 24.8$ is not an integer. Thus, we round up to the 25th position and the percentile is 67.0.

(d) 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 $32/40 = 0.80$.

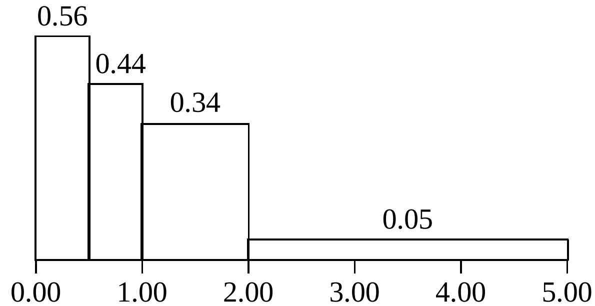
(e) Changing 14.1 to 64.1 increases the total of the 40 numbers by 50. Thus, the mean grows by $50/n = 50/40 = 1.25$. The new mean is 64.75.

The new Q_1 is $(57.6 + 58.2)/2 = 57.9$. The new median is $(64.1 + 64.3)/2 = 64.2$.

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

Class Interval	Freq.	Rel. Fr.	Density (rounded)
0.00–0.50	14	0.28	0.56
0.50–1.00	11	0.22	0.44
1.00–2.00	17	0.34	0.34
2.00–5.00	8	0.16	0.05

The density scale histogram is below.



(b) The median is $(0.88 + 1.00)/2 = 0.94$.

(c) For the 42nd percentile: $0.42(50) = 21$ is an integer. Thus, the 42nd percentile is the mean of the numbers in positions 21 and 22:

$$(0.72 + 0.73)/2 = 0.725.$$

For the 77th percentile: $0.77(50) = 38.5$ is not an integer. Thus, we round up to the 39th position and the percentile is 1.80.

(d) The one sd interval ranges from $1.20 - 1.09 = 0.11$ to $1.20 + 1.09 = 2.29$. This interval contains 38 observations; thus, the proportion within it is $38/50 = 0.76$.

(e) The deletions leave $n = 48$ numbers, a multiple of 4. Thus,

$$Q_1 = (0.39 + 0.44)/2 = 0.415 \text{ and}$$

$$Q_3 = (1.57 + 1.60)/2 = 1.585.$$

For the mean, do not sum the 48 numbers! Instead, the total of the 50 numbers is $50(1.20) = 60$; after deletions, the total of the 48 numbers is $60 - 4.05 - 4.52 = 51.43$. Thus, the mean of the 48 numbers is $51.43/48 = 1.07$.

15. (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 $[530, 671]$.

16. (a) First, $n = 18$; thus, there are 17 degrees of freedom making $t = 2.567$. The CI is

$$75.79 \pm 2.567(8.76)/\sqrt{18} =$$

$$75.79 \pm 5.30 = [70.49, 81.09].$$

(b) Note that the data are not sorted. B/c $n = 8$ is smaller than 21, we can use the exact CI. From Table A.7 the possible answers are:

- 99.2%: $[67.6, 83.6]$, and
- 93.0%: $[67.9, 78.7]$.

(c) B/c $n = 26$ is larger than 20, we must use the approximate CI. First,

$$k' = \frac{26+1}{2} - \frac{1.282\sqrt{26}}{2} =$$

$$13.5 - 3.27 = 10.23.$$

Thus, $k = 10$ and the CI is $[72.1, 77.2]$.

17. (a) The key is to save the correct intervals for the last. One interval too large means that one lower bound is too large; this must be the 58. Thus, $\mu < 58$. The one interval too small means that one upper bound is too small; this must be the 52. Thus, $\mu > 52$. The remaining intervals must be correct. Thus, $37 \leq \mu \leq 62$.

All three of these conditions must be met, so

$$52 < \mu < 58.$$

(b) Note that the first and last intervals in the list do not overlap, so they cannot both be correct. Thus, the three correct intervals Nature refers to are either: the first three or the last three. If the first three:

$$37 \leq \mu \leq 52;$$

if the last three:

$$58 \leq \mu \leq 62.$$

So the answer is that μ is in the interval $[37, 52]$ or the interval $[58, 62]$.

18. (a) The key is to save the correct intervals for the last. One interval too large means that one lower bound is too large; this must be the 86. Thus, $\mu < 86$. The one interval too small means that one upper bound is too small; this must be the 82. Thus, $\mu > 82$. The remaining intervals must be correct. Thus, $57 \leq \mu \leq 92$.

All three of these conditions must be met, so

$$82 < \mu < 86.$$

(b) Note that the first and last intervals in the list do not overlap, so they cannot both be correct. Thus, the three correct intervals Nature refers to are either: the first three or the last three. If the first three:

$$57 \leq \mu \leq 82;$$

if the last three:

$$86 \leq \mu \leq 92.$$

So the answer is that μ is in the interval $[57, 82]$ or the interval $[86, 92]$.

19. (a) Nine lower bounds are larger than 300, making nine intervals too large. One upper bound is smaller than 300, making one interval too small. Thus, $18 - 9 - 1 = 8$ intervals are correct.

(b) This means that exactly two of the lower bounds are too large. The two are 393 and 409. Thus, $\mu < 393$. But the 'exactly' tells us that the lower bound 375 is *not* too large, so $\mu \geq 375$. Putting these together, we get

$$375 \leq \mu < 393.$$

(c) This means that at least five of the upper bounds are too small. Thus, 351 is too small; the answer is

$$\mu > 351.$$

(d) By trial and error, the largest possible number of correct intervals is 11. This will happen if:

$$322 \leq \mu \leq 326 \text{ or } 344 \leq \mu \leq 345.$$

(Yes, this is a difficult question.)

20. (a) Thirteen lower bounds are larger than 600, making 13 intervals too large. Six upper bounds are smaller than 600, making six intervals too small. Thus, $24 - 13 - 6 = 5$ intervals are correct.

(b) This means that exactly three of the upper bounds are too small. The three are 518, 535 and 565. Thus, $\mu > 565$. But the 'exactly' tells us that the upper bound 568 is *not* too small, so $\mu \leq 568$. Putting these together, we get

$$565 < \mu \leq 568.$$

(c) This means that at least four of the lower bounds are too large. Thus, 720 is too large; the answer is

$$\mu < 720.$$

(d) By trial and error, the largest possible number of correct intervals is 11. This will happen if: $649 \leq \mu \leq 655$. (Yes, this is a difficult question.)

21. (a) First,

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

Thus, $s_p = 8.773$.

(b) The degrees of freedom are $16 + 9 = 25$, giving $t = 1.708$. The CI is:

$$8.00 \pm 1.708(8.773)\sqrt{1/17 + 1/10} =$$
$$8.00 \pm 5.97 = [2.03, 13.97].$$

22. (a) First,

$$s_p^2 = \frac{17(3.75)^2 + 5(8.50)^2}{17 + 5} = 27.29.$$

Thus, $s_p = 5.224$.

(b) The degrees of freedom are $17 + 5 = 22$, giving $t = 2.508$. The CI is:

$$6.25 \pm 2.508(5.224)\sqrt{1/18 + 1/6} =$$
$$6.25 \pm 6.18 = [0.07, 12.43].$$

23. (a) $0.05 <$ P-value < 0.10.

(b) P-value < 0.005.

(c) P-value $= 0.20$.

(d) P-value > 0.50.

24. (a) P-value > 0.25.

(b) P-value $= 0.05$.

(c) $0.05 <$ P-value < 0.10.

(d) P-value < 0.010.

25. The CI is $10.000 \pm h$, with $h = 4.000$. But, also

$$h = 1.345 s_p \sqrt{1/10 + 1/6}.$$

For the 95% CI,

$$h = 2.145 s_p \sqrt{1/10 + 1/6}.$$

This second h is $2.145/1.345 = 1.595$ times as large as the first h; thus, it equals $1.595(4.000) = 6.380$. Thus, the 95% CI is $10.000 \pm 6.380 = [3.620, 16.380]$.

26. The CI is $14.000 \pm h$, with $h = 4.500$. But, also

$$h = 1.711 s_p \sqrt{1/14 + 1/12}.$$

For the 99% CI,

$$h = 2.797 s_p \sqrt{1/14 + 1/12}.$$

This second h is $2.797/1.711 = 1.635$ times as large as the first h; thus, it equals $1.635(4.500) = 7.358$. Thus, the 95% CI is $14.000 \pm 7.358 = [6.642, 21.358]$.

27. (a) $740/860 = 0.860$.
   (b) $108/860 = 0.126$.
   (c) $12/120 = 0.100$.
   (d) $(12 + 42)/860 = 54/860 = 0.063$.
   (e) $698/(698 + 108) = 698/806 = 0.866$.

28. (a) $578/2909 = 0.199$.
   (b) $78/2909 = 0.027$.
   (c) $2253/2331 = 0.967$.
   (d) $156/2409 = 0.065$.
   (e) $(422 + 2253)/2909 = 2675/2909 = 0.920$.
   (f) $78/(78 + 156) = 78/234 = 0.333$.
   (g) $(422 + 78 + 156)/2909 = 656/2909 = 0.226$.

29. Let A denote female and B denote a parent.

   (a) We get the following table:

	B	B^c	Total
A	768	192	960
A^c	216	324	540
Total	984	516	1500

   (This table is acceptable only b/c I previously defined A and B; without those definitions, this table would lose points.)

   (b) $768/984 = 0.780$.

30. Let A denote that I liked the movie and B denote that RE recommended it.

   (a) We get the following table:

	B	B^c	Total
A	198	162	360
A^c	24	216	240
Total	222	378	600

   (This table is acceptable only b/c I previously defined A and B; without those definitions, this table would lose points.) $222/600 = 0.37$; i.e. 37% of the movies were recommended by RE.

   (b) $198/222 = 0.892$.
   (c) $162 + 24 = 186$.

31. (a) Her predicted value is

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

   (b) Her residual is

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

   (c) His x satisfies

$$56 = 32 + 0.4x,$$

   which yields $x = 60$.

32. (a) Her predicted value is

$$\hat{y} = 52 - 1.9(10) = 33.$$

   (b) Her residual is

$$e = 42 - 33 = 9.$$

   (c) His x satisfies

$$14 = 52 - 1.9x,$$

   which yields $x = 20$.

33. 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

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

34. The second midterm is X and the first midterm is Y. Thus,

$$\bar{x} = 70.0, s_X = 10.00, \bar{y} = 60.0,$$

$$s_Y = 15.00, \text{ and } r = 0.70.$$

Thus,

$$b_1 = 0.70(15/10) = 1.05, \text{ and}$$

$$b_0 = 60 - 1.05(70) = -13.5.$$

The regression line is

$$\hat{y} = -13.5 + 1.05x.$$

35. 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

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

36. Height is X and weight is Y. For Frank, $x = 66$ and $y = 160$. Also, $\bar{x} = 68$ and $\bar{y} = 160$. Finally, Franks's residual Tom's line is $e = +6$. 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,

$$160 = b_0 + 68b_1.$$

Using information on Frank,

$$154 = b_0 + 66b_1.$$

Subtracting these, we get

$$2b_1 = 6, \text{ or } b_1 = 3.$$

Thus,

$$160 = b_0 + 204, \text{ or } b_0 = -44.$$

Thus, the regression line is

$$\hat{y} = -44 + 3x.$$