Practice Exam Questions and Solutions
for the Final Exam
Statistics 301, Professor Wardrop

Chapter 12

1. 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	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) 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?

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

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

   (f) Refer to (c). Ralph learns that the 14.1 is a misprint. The actual observation is 54.1. Recalculate the mean after changing the 14.1 to 54.1.

Chapter 15

2. 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.

Use the following information to answer questions 3–6. 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 7.

3. Given that $\mu = 300$, how many of the confidence intervals are correct?

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

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

6. 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?

7. 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.**

Chapter 16

8. 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

   Calculate s_p.

9. 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.

Chapter 8

10. 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 either: a ratio, like 72/312 or 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?

11. (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.

Chapter 13

12. 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.

13. 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.

14. 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.

15. A regression analysis is performed with $n = 4$ observations. You are given the following information.

x	e
0	e_0
2	e_2
3	$+3$
5	-2

Calculate the values of e_0 and e_2. Be sure to label your answers so that I can tell which is which. (**Hint:** Remember that $\sum e = 0$ and $\sum (xe) = 0$.)

16. A regression analysis yields the following regression line:

$$\hat{y} = 20 + 6x.$$

Which **one or more** of the following five sets of statistics can be for these data?

(a) $s_Y = 20$, $s_X = 5$, $\bar{y} = 110$ and $\bar{x} = 15$.

(b) $s_Y = 40$, $s_X = 5$, $\bar{y} = 110$ and $\bar{x} = 15$.

(c) $s_Y = 40$, $s_X = 5$, $\bar{y} = 110$ and $\bar{x} = 25$.

(d) $s_Y = 20$, $s_X = 5$, $\bar{y} = 110$ and $\bar{x} = 25$.

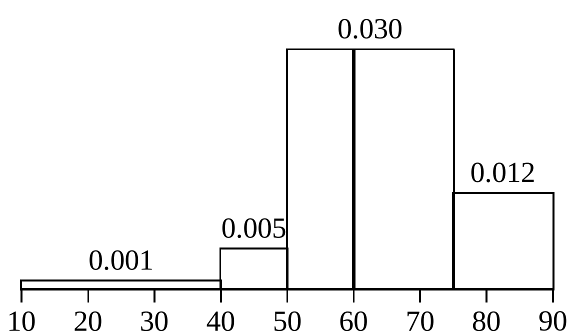
(e) $s_Y = 50$, $s_X = 3$, $\bar{y} = 110$ and $\bar{x} = 15$.

Solutions

1. (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 the mean of the numbers in positions 20 and 21; i.e. 63.0 and 64.3—the median is 63.65. The lower half of the data has 20 observations. The first quartile is the mean of 55.4 and 57.6; i.e. 56.5.

(c) 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 $32/40 = 0.80$.

(d) The median is the number in position 20; i.e. 64.3. The lower half of the data has 19 observations. The first quartile is 57.6.

(e) The total of the 40 numbers is 40 times the mean, or $40(63.50) = 2540$. After deleting 14.1, the total of the 39 remaining numbers is $2540 - 14.1 = 2525.9$. The mean of the 39 remaining numbers is $2525.9/39 = 64.77$.

(f) Changing 14.1 to 54.1 increases the total of the 40 numbers by 40. Thus, the mean grows by $40/n = 40/40 = 1$. The new mean is 64.50.

2. (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 $[530, 671]$.

3. 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.

4. 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

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

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

$$\mu > 351.$$

6. 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.)

7. (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 $[58, 62]$.

8. First,

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

   Thus, $s_p = 8.773$.

9. (a) $0.05 <$ P-value < 0.10.
   (b) P-value < 0.005.
   (c) P-value $= 0.20$.

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

11. 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) $768/984 = 0.780$.

12. (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 $x = 60$.

13. 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

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

14. 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

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

15. The fact that the e's sum to zero implies:

$$e_0 + e_2 = -1.$$

The fact that the xe's sum to zero implies:

$$2e_2 = 1.$$

Thus, $e_2 = 0.5$ and $e_0 = -1.5$.

16. There are no r's in this list, so that should alert you how to do the problem. The slope, 6, must equal

$$rs_Y/s_X.$$

In (a) and (d), $s_Y/s_X = 4$. Thus, remembering $-1 \leq r \leq +1$, the slope must be between -4 and $+4$. Thus, (a) and (d) are wrong.

Next,

$$b_0 = \bar{y} - b_1\bar{x} = 20.$$

This is true for (b) and (e), but not (c). Thus, (b) and (e) could be for this regression line.