Practice Exam Questions and Solutions for the Final Exam; Fall, 2008
Statistics 301, Professor Wardrop
Part B, Chapters 16, 8 and 13

Chapter 16

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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

7. 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?

8. 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?

9. (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.

10. (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

11. 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.

12. 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.

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. 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.

15. 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.

16. (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.

17. 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 $\sum(xe) = 0$.)

18. A regression analysis is performed with $n = 5$ observations. You are given the following information.

x	e
0	-1
1	$+2$
2	-2
3	e_3
4	e_4

Calculate the values of e_3 and e_4. Be sure to label your answers so that I can tell which is which. (**Hint:** Remember that $\sum e = 0$ and $\sum(xe) = 0$.)

Solutions

1. (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:

$$8.00 \pm 1.708(8.773)\sqrt{1/17 + 1/10} =$$

$$8.00 \pm 5.97 = [2.03, 13.97].$$

2. (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:

$$6.25 \pm 2.508(5.224)\sqrt{1/18 + 1/6} =$$

$$6.25 \pm 6.18 = [0.07, 12.43].$$

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

(b) P-value < 0.005.

(c) P-value $= 0.20$.

(d) P-value > 0.50.

4. (a) P-value > 0.25.

(b) P-value $= 0.05$.

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

(d) P-value < 0.010.

5. 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 $10.000 \pm 6.380 = [3.620, 16.380]$.

6. 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 $14.000 \pm 7.358 = [6.642, 21.358]$.

7. (a) $740/860 = 0.860$.

(b) $108/860 = 0.126$.

(c) $12/120 = 0.100$.

(d) $(12 + 42)/860 = 54/860 = 0.063$.

(e) $698/(698 + 108) = 698/806 = 0.866$.

8. (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) $(422 + 78 + 156)/2909 = 656/2909 = 0.226$.

9. 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) $768/984 = 0.780$.

10. 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) $162 + 24 = 186$.

11. (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 $x = 60$.

12. (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 $x = 20$.

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. 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

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

15. 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

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

16. 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

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

17. 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 $e_0 = -1.5$.

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

$$e_3 + e_4 = 1.$$

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

$$3e_3 + 4e_4 = 2.$$

Thus, $e_4 = -1$ and $e_3 = 2$.