

## Exam II

No books, notes or electronic devices may be used on this exam.

1. (35 points) In a certain class, midterm scores average out to 75 with an SD of 15. Scores on the final average out to 70 with an SD of 15. The correlation between midterm scores and final scores is about 0.50.

a) Write down the equation for the regression line to predict final exam score from midterm score.

b) Sketch the regression line.

c) Estimate the final exam scores for students whose midterm scores were 85, 45 and 75.

d) Estimate the midterm score for a student who scored 90 on the final exam.

2. (15 points)

a) For a representative sample of cars, would the correlation between the age of the car and its gasoline economy (miles per gallon) be positive or negative? Explain

b ) The correlation between gasoline economy and income of owner turns out to be positive? How do you account for this association?

3. (10 points) Test scores on a certain exam average out to 75 with a standard deviation of 10. One student scored 84. Was that a very high score? Moderately high? Or about average? Justify your answer.

(15 points)

4. Consider the following data.

$x$	$y$
0	2
1	4
2	3

a) Make a scatter plot of the data.

b) An investigator uses the line  $y = 2x + 1$  to predict  $y$  from  $x$ . Add the line to your sketch.

c) Calculate the RMS error for the line:  $\sqrt{\frac{1}{n} \sum (\text{predicted } y - \text{actual } y)^2}$ .

5. (10 points) For women age 25 and over in the U.S. in 2005, the relationship between age and educational level (years of schooling completed) can be summarized as follows.

average age  $\approx 50$  years,                      SD  $\approx 16$  years  
average education level  $\approx 13.2$  years,                      SD  $\approx 3.0$  years,                       $r \approx -0.2$

a) True or false: For women age 25 and older in the U.S. in 2005, older women tended to be less educated than younger ones.

b) True or false, and explain: as you get older, you become less educated. If this statement is false, what could account for the negative correlation?

6. (15 points) Suppose that the regression line of  $y$  on  $x$  is

$$y = \left( \frac{\sqrt{3}}{2} \right) \left( \frac{8}{\sqrt{3}} \right) (x - 1)$$

where  $\text{mean}_x = 1$ ,  $\text{mean}_y = 0$ ,  $\text{SD}_x = \sqrt{3}$ ,  $\text{SD}_y = 8$  and  $r = \sqrt{3}/2$ .

- a) Calculate the RMS error for regression  $\text{RMS} = \text{SD}_y \sqrt{1 - r^2}$ . What can you say about the RMS for the SD line?

- b) If  $x = 4$ , then we expect 95% of the  $y$ -values to be between \_\_\_\_\_ and \_\_\_\_\_ .

## Exam II

No books, notes or electronic devices may be used on this exam.

1. (35 points) In a certain class, midterm scores average out to 70 with an SD of 15. Scores on the final average out to 75 with an SD of 15. The correlation between midterm scores and final scores is about 0.50.

a) Write down the equation for the regression line to predict final exam score from midterm score.

b) Sketch the regression line.

c) Estimate the final exam scores for students whose midterm scores were 80, 40 and 70.

d) Estimate the midterm score for a student who scored 95 on the final exam.

2. (15 points)

a) For a representative sample of cars, would the correlation between the age of the car and its gasoline economy (miles per gallon) be positive or negative? Explain

b ) The correlation between gasoline economy and income of owner turns out to be positive? How do you account for this association?

3. (10 points) Test scores on a certain exam average out to 75 with a standard deviation of 3. One student scored 84. Was that a very high score? Moderately high? Or about average? Justify your answer.

(15 points)

4. Consider the following data.

$x$	$y$
0	2
1	4
2	3

a) Make a scatter plot of the data.

b) An investigator uses the line  $y = 2x + 1$  to predict  $y$  from  $x$ . Add the line to your sketch.

c) Calculate the RMS error for the line:  $\sqrt{\frac{1}{n} \sum (\text{predicted } y - \text{actual } y)^2}$ .

5. (10 points) For women age 25 and over in the U.S. in 2005, the relationship between age and educational level (years of schooling completed) can be summarized as follows.

average age  $\approx 50$  years,                      SD  $\approx 16$  years  
average education level  $\approx 13.2$  years,                      SD  $\approx 3.0$  years,                       $r \approx -0.2$

a) True or false: For women age 25 and older in the U.S. in 2005, older women tended to be less educated than younger ones.

b) True or false, and explain: as you get older, you become less educated. If this statement is false, what could account for the negative correlation?

6. (15 points) Suppose that the regression line of  $y$  on  $x$  is

$$y = \left( \frac{\sqrt{3}}{2} \right) \left( \frac{8}{\sqrt{3}} \right) (x - 1)$$

where  $\text{mean}_x = 1$ ,  $\text{mean}_y = 0$ ,  $\text{SD}_x = \sqrt{3}$ ,  $\text{SD}_y = 8$  and  $r = \sqrt{3}/2$ .

- a) Calculate the RMS error for regression  $\text{RMS} = \text{SD}_y \sqrt{1 - r^2}$ . What can you say about the RMS for the SD line?

- b) If  $x = 4$ , then we expect 68% of the  $y$ -values to be between \_\_\_\_\_ and \_\_\_\_\_ .