

SECTION II

PART A

Questions 1–5

Spend about 65 minutes on this part of the exam.
Percentage of Section II grade—75

You must show all work and indicate the methods you use. You will be graded on the correctness of your methods and on the accuracy of your results and explanations.

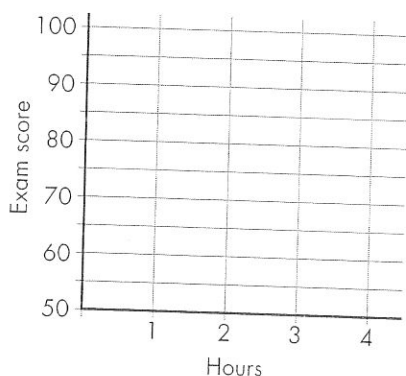
1. A student is interested in estimating the average length of words in a 600-page textbook, and plans the following three-stage sampling procedure:
 - (1) Noting that each of the 12 chapters has a different author, the student decides to obtain a sample of words from each chapter.
 - (2) Each chapter is approximately 50 pages long. The student uses a random number generator to pick three pages from each chapter.
 - (3) On each chosen page the student notes the length of every tenth word.
 - (a) The first stage above represents what kind of sampling procedure? Give an advantage in using it in this context.
 - (b) The second stage above represents what kind of sampling procedure? Give an advantage in using it in this context.
 - (c) The third stage above represents what kind of sampling procedure? Give a *disadvantage* to using it dependent upon an author's writing style.
2. (a) Suppose that nationwide, 69% of all registered voters would answer "Yes" to the question "Do you consider yourself highly focused on this year's presidential election?" Which of the following is more likely: an SRS of 50 registered voters having over 75% answer "Yes," or an SRS of 100 registered voters having over 75% answer "Yes" to the given question? Explain.
 - (b) A particular company with 95 employees wishes to survey their employees with regard to interest in the presidential election. They pick an SRS of 30 employees and ask "Do you consider yourself highly focused on this year's presidential election?" Suppose that in fact 60% of all 95 employees would have answered "Yes." Explain why it is not reasonable to say that the distribution for the count in the sample who say "Yes" is a binomial with $n = 30$ and $p = .6$.

- (c) Suppose that nationwide, 78% of all registered voters would answer "Yes" to the question "Do you consider yourself highly focused on this year's presidential election?" You plan to interview an SRS of 20 registered voters. Explain why it is not reasonable to say that the distribution for the proportion in the sample who say "Yes" is approximately a normal distribution.
3. An instructor takes an anonymous survey and notes exam score, hours studied, and gender for the first exam in a large college statistics class ($n = 250$). A resulting regression model is:

$$\text{Score} = 50.90 + 9.45(\text{Hours}) + 4.40(\text{Gender})$$

where *Gender* takes the value 0 for males and 1 for females.

- (a) Provide an interpretation in context for each of the three numbers appearing in the above model formula.
- (b) Sketch the separate prediction lines for males and females resulting from using 0 or 1 in the above model.



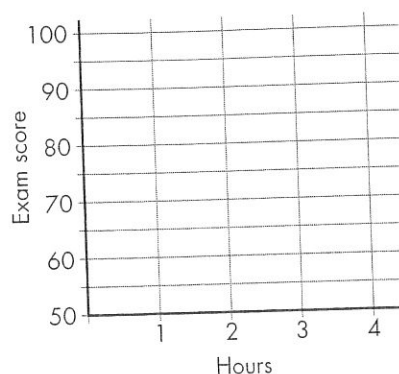
Looking at the data separately by gender results in the following two regression models:

For males: $\text{Score} = 51.8 + 8.5(\text{Hours})$

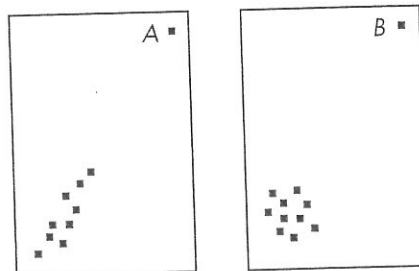
For females: $\text{Score} = 54.4 + 10.4(\text{Hours})$

- (c) What comparative information do the slope coefficients from these two models give that does not show in the original model?

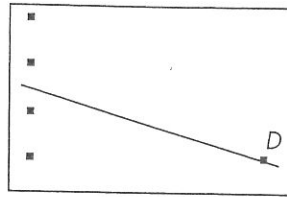
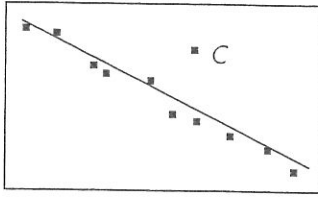
- (d) Sketch the prediction lines given by these last two models.



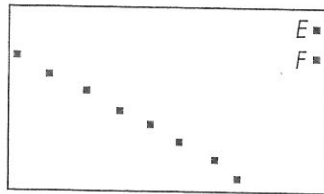
4. A laboratory is testing the concentration level in milligrams per milliliter for the active ingredient found in a pharmaceutical product. In a random sample of five vials of the product, the concentrations were measured at 2.46, 2.57, 2.70, 2.64, and 2.54 mg/ml.
- Determine a 95% confidence interval for the mean concentration level in milligrams per milliliter for the active ingredient found in this pharmaceutical product.
 - Explain in words what effect an increase in confidence level would have on the width of the confidence interval.
 - Suppose a concentration above 2.70 milligrams per milliliter is considered dangerous. What conclusion is justified by your answers to (a) and (b)?
5. A point is said to have *high leverage* if it is an outlier in the x -direction, and a point is said to be *influential* if its removal sharply changes the regression line.
- In the scatterplots below, compare points A and B with regard to having high leverage and with regard to being influential.



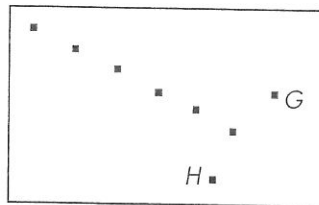
- (b) In the scatterplots below, compare points C and D with regard to residuals and influence on β_1 , the slope of the regression line.



- (c) In the scatterplot below, compare the effect of removing point E or point F to that of removing both E and F .



- (d) In the scatterplot below, compare the effect of removing point G or point H to that of removing both G and H .



SECTION II

PART B

Question 6

Spend about 25 minutes on this part of the exam.

Percentage of Section II grade—25

6. In two random samples of adults, one from each of two communities, counts were made of the number of people with different years of schooling and tabulated as follows:

Years of schooling	Community A	Community B
8	20	15
10	35	40
12	85	40
14	100	20
16	90	40
18	40	55
20	20	25

$$n = 390$$

$$\bar{x} = 14.08$$

$$s = 2.93$$

$$n = 235$$

$$\bar{x} = 14.51$$

$$s = 3.71$$

- (a) Is there evidence of a difference in the mean years of schooling between the two communities?
- (b) Is there evidence of a difference in the distribution of years of schooling in the two communities?
- (c) Use histograms to help explain the above results.



If there is still time remaining, you may review your answers.