Statistics 257 Midterm – October 21, 2005

This exam has 2 problems and 6 numbered pages.

You have 50 minutes to complete this exam. Please read all instructions carefully, and check your answers. Show all work neatly and in order, and clearly indicate your final answers. Answers must be justified whenever possible in order to earn full credit. **Unless otherwise specified, no credit will be given for unsupported answers, even if your final answer is correct.** *Points will be deducted for incoherent, incorrect, and/or irrelevant statements.*

Calculators are permitted, as well as an $8\frac{1}{2} \times 11$ double-sided page of handwritten notes. No other aids are allowed.

You must answer all of the questions in the space provided. Note that blank space is NOT an indication of a question's difficulty.

Name: _____

Instructor: Michael Kozdron

Problem	Score
1	
2	

TOTAL: _____

1. (*25 points*) Suppose that you are interested in estimating the average amount of money that was spent on books for Fall 2005 classes by full-time undergraduate students at the University of Regina.

Explain how you would design a survey which uses regression (or ratio) estimation to help you estimate this average amount of money spent on books.

Be sure to explicitly state your population of interest, an appropriate frame/list, and appropriate sampling units, as well as your variable of interest, and your auxillary variable.

Remark: The solution to Problem #2 does not use your answer to Problem #1.

2. (*25 points*) After designing a survey to estimate the average amount of money that was spent on books for Fall 2005 classes by University of Regina full-time undergraduate students, you conducted a simple random sample of $n = 50$ students from among the 12 700 full-time undergraduates.

Let Y_i denote the amount spent by the ith student on books, so that the parameter $\overline{Y}$ denotes the average amount spent on books by U of R undergraduates. It is the parameter $\overline{Y}$ you are interested in estimating.

Furthermore, suppose that X is the auxillary variable for which it is known that $\overline{X} = 108$.

You observed data (x_i, y_i) for each sampled student, $i = 1, \ldots, 50$, and calculated

$$\sum_{i=1}^{50} y_i = 42\ 500, \quad \sum_{i=1}^{50} y_i^2 = 37\ 250\ 000, \quad \sum_{i=1}^{50} x_i = 5000, \quad \sum_{i=1}^{50} x_i^2 = 516\ 500,$$

$$\sum_{i=1}^{50} x_i y_i = 4\ 351\ 000.$$

(a) Construct an approximate 95% confidence interval for $\overline{Y}$ based on the values of y alone.

Recall that

$$\sum_{i=1}^{50} y_i = 42\ 500, \quad \sum_{i=1}^{50} y_i^2 = 37\ 250\ 000, \quad \sum_{i=1}^{50} x_i = 5000, \quad \sum_{i=1}^{50} x_i^2 = 516\ 500,$$

$$\sum_{i=1}^{50} x_i y_i = 4\ 351\ 000.$$

(b) Construct an approximate 95% confidence interval for $\overline{Y}$ using regression estimation.

(continued)

4

(c) Compute the relative efficiency of the regression estimator you found in **(b)** to the simple random sampling estimator you found in **(a)**. Which estimator, if either, do you conclude is preferrable for estimating the average amount spent on books?

(d) Suppose that you want to be 95% confident that a simple random sampling estimate of $\overline{Y}$ is accurate within \$20. In this case, how many students should be included in the sample?

(The End.)