

Make sure that this examination has 12 numbered pages

University of Regina
Department of Mathematics & Statistics
Final Examination
200530
(December 16, 2005)

Statistics 257-001
Applied Sampling Techniques

Name: _____ Student Number: _____

Instructor: Michael Kozdron

Time: 3 hours

Read all of the following information before starting the exam.

*You have 3 hours 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.** Several problems require written explanations in context. Only complete solutions written in the context specified by the problem will be awarded full points, and points will be deducted for incoherent, incorrect, and/or irrelevant statements.*

You may use standard notation; however, any new notations or abbreviations that you introduce must be clearly defined.

*Calculators are permitted; however, you must still show all your work. You are also permitted to have **TWO** 8.5×11 pages of handwritten notes (each double-sided) for your personal use. Other than these exceptions, no other aids are allowed.*

*Throughout this exam, you may perform calculations accurate to **three** decimal places.*

Note that blank space is not an indication of a question's difficulty. The order of the test questions is essentially random; they are not intentionally written easiest-to-hardest.

*This test has **12** numbered pages with **8** questions totalling **145** points. The number of points per question is indicated.*

Solutions will be available on the Statistics 257 course web site after the examinations have all been graded.

DO NOT WRITE BELOW THIS LINE

Problem 1 _____ Problem 2 _____ Problem 3 _____
Problem 4 _____ Problem 5 _____ Problem 6 _____
Problem 7 _____ Problem 8 _____

TOTAL _____

1. (18 points) Cypress Hills interprovincial park is a 400 km² prairie parkland that straddles southwestern Saskatchewan and southeastern Alberta. The pine and spruce forests, and lush grasslands are filled with hundreds of different species of birds and animals.

Among the animals inhabiting Cypress Hills is the *pronghorn* which, according to the National Wildlife Federation is the fastest land animal in the western hemisphere, and second in the world to the cheetah. In 1915, after the North American pronghorn was almost hunted to extinction, it was estimated that there were only 15 000 animals worldwide. Even though there are roughly 700 000 pronghorns today, some subspecies are still on the brink of extinction.

A University of Regina biologist is interested in studying the pronghorns of Cypress Hills, and he manages to capture and tag 59 pronghorns. He returns to Cypress Hills one year later, and captures 45 pronghorns, of which 9 had been previously tagged.

(a) Compute an approximate 95% confidence interval for the true total number of pronghorns in Cypress Hills using the Petersen estimator.

(b) Compute an approximate 95% confidence interval for the true total number of pronghorns in Cypress Hills using the Chapman estimator.

According to the Saskatchewan Park Office, a species of bird that is very commonly seen in Cypress Hills is the *yellow-bellied sapsucker*.

A University of Regina ornithologist is interested in studying the yellow-bellied sapsucker of Cypress Hills. She randomly selects a path 30 km long through the parkland, and counts 258 yellow-bellied sapsuckers within 100 metres of her path.

(c) Based on this data, estimate the total number of yellow-bellied sapsuckers in Cypress Hills. (Watch your units!)

2. (7 points) A senior administrator at the University of Regina wishes to estimate the proportion of its undergraduate students that have plagiarized on final exams, a sensitive subject. Since there was some concern that students might be unwilling to disclose their plagiarisms to a university official, the following *random response method* was used. The university official constructs a deck of 50 cards. On 38 of them are marked **N** for *never plagiarized on a final exam* and 12 of them are marked **P** for *have plagiarized on a final exam at least once*. Each sampled student was asked to draw a card from the deck and to respond *yes* if the letter agrees with the group that student belongs to. The results are as follows:

total number of students	number sampled	number answering yes
8972	280	198

Construct a 95% confidence interval for the proportion of undergraduates who have plagiarized on a final exam at least once.

3. (10 points) The manager of the University of Regina's *Physical Plant Operations* wishes to estimate the average amount of time (in minutes) a snowplow operator needs to clear the sidewalks in front of the College West building during the winter. There are 62 snowplow operators that work for Physical Plant Operations, of which eight operators are selected at random and timed. The observed results (time, in minutes) are shown in the table below.

12		13
11		9
17		13
12		9

- (a) Construct an approximate 95% confidence interval for the average amount of time taken to plow the sidewalks by all Physical Plant snowplow operators.

- (b) Physical Plant also employs lawn mower operators to cut the grass in front of the College West building in the summer. The manager of Physical Plant found that an approximate 95% confidence interval for the average amount of time taken to cut the grass by all Physical Plant lawn mower operators is 10 ± 1 minutes. Based on your answer to (a), is there a significant difference in the amount of time spent cutting the grass in the summer versus plowing the sidewalks in the winter? Explain.

4. (20 points) A statistician is interested in estimating the average amount of money that was spent on required textbooks for Fall 2005 classes by full-time undergraduate students at the University of Regina. A simple random sample of $n = 100$ full-time undergraduates was conducted. Each undergraduate sampled was enrolled in 15 credit hours for Fall 2005, and was asked the following question:

“Approximately, how much did you spend in total on **required** textbooks and manuals for your Fall 2005 classes?”

The respondents were asked to select their answer from among a range of possible dollar amounts. The data collected is summarized below:

0–100	101–200	201–300	301–400	401–500	501–600
3	17	7	28	31	14

Based on this data, compute a reasonable point estimator for the average amount of money spent by full-time (enrolled in 15 credit hours) undergraduate students on required textbooks for Fall 2005 classes.

Note: Part of your solution **must** include a brief justification for why you believe that your estimator is reasonable. You do **NOT** need to compute the estimated variance for your estimator.

5. (18 points) The *Canadian Association of University Teachers (CAUT)* is interested in determining the average annual salary of full-time faculty members at the University of Regina. According to its *Facts & Figures* booklet, the university employs 453 full-time faculty members. Unfortunately for the CAUT, it is extremely difficult to find published salaries. It is, however, possible to determine the total number of years that any faculty member has been employed by the university. That is, if Y denotes salary (in thousands (\$1000) of dollars), and X denotes years employed, then

$$X_T = \sum_{i=1}^{453} X_i = 5436.$$

CAUT decides to randomly survey 12 faculty members, and collects the following data, where y_i denotes the salary (in thousands (\$1000) of dollars) and x_i denotes the number of years worked by the i th faculty member sampled:

$$\sum_{i=1}^{12} y_i = 897.6, \quad \sum_{i=1}^{12} y_i^2 = 73,560, \quad \sum_{i=1}^{12} x_i = 132, \quad \sum_{i=1}^{12} x_i^2 = 1473, \quad \sum_{i=1}^{12} x_i y_i = 9820.$$

Find an approximate 95% confidence interval for \bar{Y} , the average annual salary of University of Regina full-time faculty members, using the estimation methods given in (a) and (b) below.

(a) *ratio estimation*:

(b) *regression estimation:*

(c) Compute the relative efficiency of these two estimators. Is there sufficient evidence to favour one over the other? Why or why not?

6. (*14 points*) A political scientist from the University of Regina has developed a test to measure the degree of awareness of current events. She wants to estimate the average score that would be achieved on this test by all students at *Central High School*. The principal of Central High School will not let the experimenter randomly select students out of classes in session, but he will allow her to interrupt a small number of classes for the purpose of giving the test to every member of the class. Thus, the experimenter decides that cluster sampling is most appropriate, and selects 5 classes at random from the 40 classes in session at a particular hour. The test is given to each member of the sampled classes, and the results are shown in the table below. It is also known that there are 850 students enrolled in Central High School.

class i	number of students in class i	average score for class i
1	21	13
2	18	14
3	22	17
4	32	11
5	17	15

(a) Compute an unbiased estimator of the average score that would be achieved by all Central High School students on this test.

(b) Compute the estimated standard error of your estimator in (a).

7. (22 points) The political scientist from Problem 6 contacted the principal at *Eastern High School* and asked to administer her test to measure the degree of awareness of current events to Eastern High School students. Since the principal of Eastern High School *will* let the experimenter randomly select students out of classes in session, the political scientist decides that a stratified random sample would be most appropriate. She decides to stratify by grade level, and obtains the following results. (Grades 9 and 10 combined are called juniors, and grades 11 and 12 combined are called seniors.)

STRATA	strata size	sample size	sample mean	sample variance
Juniors (Grades 9, 10)	350	45	12	4
Seniors (Grades 11, 12)	450	55	16	9

(a) Find an approximate 95% confidence interval for \bar{Y} , the average score that would be achieved by all Eastern High School students on this test.

Suppose that the political scientist had wanted the variance for the estimation of \bar{Y} to be $V = \text{Var}(\bar{y}_{ST}) = 0.02$. What sample size should she have used and how should it have been allocated for each of the allocation methods given in **(b)** and **(c)** below?

(b) *proportional allocation:*

(c) *Neyman allocation:*

8. (*36 points*) Skytrax, an independent research firm, conducts an annual world survey of airline passengers, and presents the World Airline Awards to deserving airlines based on this survey. The survey focuses on 35 different aspects of passenger satisfaction such as the user-friendliness of the internet site, cabin service and comfort, and the friendliness, attentiveness and language skills of the staff. From its 2005 survey results, Skytrax judged Air Canada to be the Best Airline in North America.

However, a consumer protection group wishes to estimate the total number of complaints that will be received by both Canadian airlines (namely, Air Canada and WestJet) in January 2006, and has hired you to help them. Air Canada has agreed to tell you the total number of passengers that will fly their airline in January 2006 and to allow you to distribute questionnaires to its passengers during this month. Unfortunately, WestJet will not reveal this information to you and will not let you survey its passengers.

Consequently, the consumer protection group would like you to extrapolate your results for Air Canada to WestJet in a meaningful way.

- (a) Identify the population of interest, the sampling frame, the sampling units, and the variable(s) of interest.

- (b) Propose a detailed method to estimate the total number of complaints that Air Canada will receive in January 2006.

(c) What are two pieces of information that you should obtain from each sampled Air Canada passenger in order to compare the Air Canada results to WestJet?

(d) With explicit reference to your answer to (c), how could you extrapolate your estimate from (b) for Air Canada to the total number of complaints received by WestJet in January 2006?