Statistics 257 Midterm – October 25, 2004

This exam has 2 problems and 8 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. A dictionary will be provided.

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. (*75 points*)

A sociologist at the University of Saskatchewan is concerned about faculty publications at her university. She has decided to conduct a survey of her colleagues and has chosen to distribute a questionnaire to them.

After preparing the questionnaire, she sent a copy of it by mail to each faculty member across campus. She also wrote an email to all university faculty members urging them to complete and return her survey. The following week she also sent a reminder by email.

Some of the data she collected is summarized below.

Department	Number of Faculty Members	Number of Respondents
Literature	51	32
Classics	33	16
Philosophy	35	18
History	55	33
Linguistics	36	21
Political Science	53	24
Sociology	52	27

Department	Total Publications among Respondents	Range in Number of Publications
Literature	80	0 to 4
Classics	72	0 to 8
Philosophy	108	1 to 9
History	132	0 to 16
Linguistics	42	1 to 5
Political Science	84	0 to 4
Sociology	81	0 to 8

(a) To analyze this survey, the sociologist decides to treat the returned questionnaires as forming a random sample. In the context of this problem, what are some of the concerns or considerations in doing so? Conversely, what are some of the advantages in doing so?
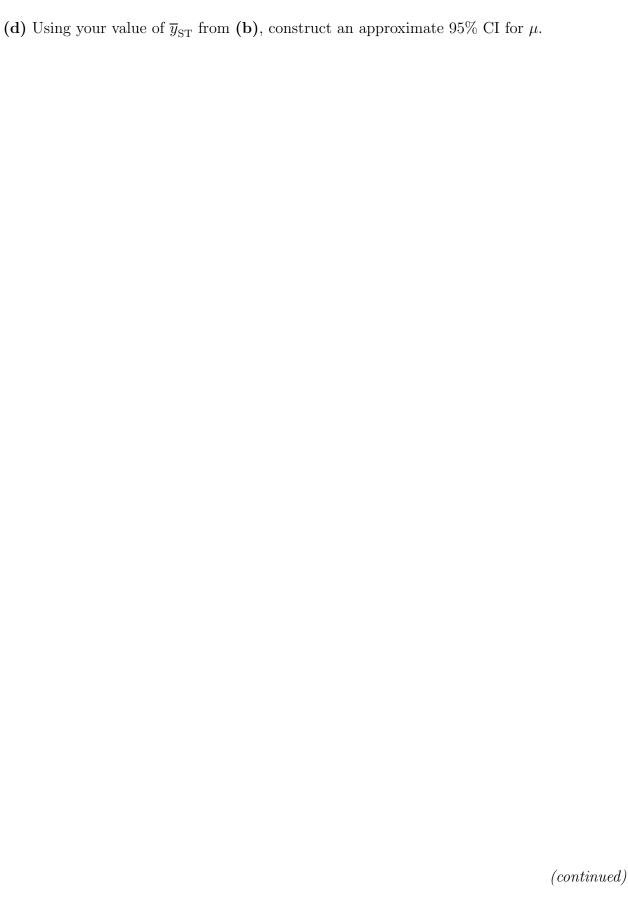
(continued)

1

In order to compare publication rates, the sociologist decides that since publication requirements vary between departments, it is natural to treat each department as a stratum. For the remaining problems, assume that the returned questionnaires in each stratum do form a simple random sample for that stratum.

(b) The sociologist is interested in μ, the average number of faculty publications across the departments. Compute $\overline{y}_{\mathrm{ST}}$ as an estimator of μ.

(c) Verify that for the stratified random sample described above, $\overline{y}_{\mathrm{ST}}$ is, in fact, an unbiased estimator of μ.

(d) Using your value of $\overline{y}_{\mathrm{ST}}$ from **(b)**, construct an approximate 95% CI for μ.

(e) Is there a statistically significant difference in the average number of publications for faculty members in Linguistics compared to History? Why or why not?

(f) The sociologist is also interested in average faculty salaries, and has decided to send a follow-up questionnaire to some of those $N = 171$ who responded to her first survey. Since all 171 faculty members who responded to her first questionnaire indicated that they were willing to be contacted again, non-response is not an issue. Using records provided by the Faculty Association, she believes that each stratum's average salary has a standard deviation of \$10000. Assuming that it will cost her \$16 to prepare and mail the questionnaire to Literature and Classics, but only \$9 to prepare and mail it to the other departments, find the approximate sample size and the necessary allocation to estimate the average salary within \$1000 (while minimizing her overall cost).

(g) Unfortunately, the sociologist only has $1122 left in her grant to spend on conducting the survey on faculty salaries described in **(f)**. Therefore, if the total cost of sampling is fixed at $1122, choose the sample size that minimizes the variance of the average faculty salary estimator for this fixed cost.

(h) The results that you have found for the sociologist here cannot be used to infer results about all University of Saskatchewan faculty members. Why is this the case? To what population can her reults be applied?

2. (*25 points*)

(a) List and briefly describe the two main steps in a SAS program.

STEP ONE:

STEP TWO:

(b) The following SAS code is used to create the SAS data set `country`.

```
DATA country;
     INFILE '/u1/stat257/Country.dat';
     INPUT hemisphere $ cont $ country $ pop92 urban gdp
           lifeexpm lifeexpf birthrat deathrat;
RUN;
```

After creating the SAS `country` data set described above, the following output was generated.

```
                        The SAS System                       13
                             11:19 Monday, October 18, 2004

                        The MEANS Procedure
                     Analysis Variable : pop92

                  N
     cont        Obs         Mean         Median         Variance
     -------------------------------------------------------------
     Africa      42      16.1790000      8.6930000      338.0928272

     Asia        19     154.0138947     42.6420000       99939.66

     CAmerica    11       5.3167273      4.9490000        9.5336912

     Europe      20      19.1538000     10.0400000      377.8315688

     MEast       13      16.4275385     10.3940000      419.2600556

     NAmerica     3     125.4306667     92.3800000       13953.57

     SAmerica    11      28.1101818     13.5280000        1986.72

     SPacific     3       8.3066667      4.0060000       64.4239003
     -------------------------------------------------------------
```

Write down what you would need to submit in SAS in order to generate this output?