Stat 257.01 Fall 2004
Assignment #11

This assignment is **never** due.

The following problems constitute a *selection* of problems that I think would be beneficial for you to do in preparation for the final exam. While this list is rather long and extensive, you **can expect** to have both problems which require written explanations, and problems which require application of formulæ. With this selection of problems I have tried to give you both types to work on. One very important note: Since these problems are intended to be given as *homework*, many of the numbers do not "work out nicely." Of course, that is reality. But on the final exam, I have endeavoured to make the calculations give exact answers whenever possible.

- 2.9, 2.20
- 3.1 (perhaps you can use a ratio/regression/difference estimator), 3.3 (Don't forget about cluster sampling!), 3.16
- 4.17, both 4.18 and 4.19, 4.22, 4.24, 4.38
- both 5.19 and 5.20, 5.22, 5.23, 5.24
- 6.4, 6.6, 6.9, 6.31
- 6.34. Compute both a regression estimate and a difference estimate, and in each case construct an approximate 95% confidence interval for the average idling time of automobiles with 2.5 litre engines. Also compute the relative efficiency of these two estimators and decide if one is preferable to the other in this case.
- 7.6, 7.8, 7.14, 7.22 (really, who wants to *eat* defective wafers :–), 7.24, 7.26 (including cluster sampling)
- 8.16, 8.17, 8.20, 8.31, 8.34, 8.35
- 11.1, 11.10
- 12.2, 12.4, 12.11, 12.12, 12.13, 12.14
- Give two reasons why systematic sampling provides a useful alternative to simple random sampling.
- Carefully state what is meant by a *1-in-k systematic sample.*
- Define *cluster sampling* as precisely as you can.
- Briefly outline the steps an experimenter might take to draw a cluster sample.
- What is a table of random digits? What is the purpose of a table of random digits in simple random sampling?
- What is the relationship of a *frame* to a *population*?

- Recall that one way (of two equivalent ways) to define the population ratio R is

$$R = \frac{\mu_y}{\mu_x}$$

where μ_y, μ_x are the population means for the y-population, x-population, respectively. Recall further that the ratio estimator r given by

$$r = \frac{\overline{y}}{\overline{x}}$$

is used to estimate R. Carefully explain why r is *not* an unbiased estimator of R. That is, r is a biased estimator of R! Although our preference is to use unbiased estimators whenever possible, the utility of the ratio estimator r is that it is easily computed and proves in practice, when it is applicable, to be quite powerful.

- Consider the following scenarios. Briefly explain why each does *not* describe a *probability sample*. In the context of the scenario, is there an alternative sampling scheme that might be more appropriate? (Remember that many factors affect the selection of a survey scheme including cost and convenience.)

  - A sample of coal is taken from the top 15–30 cm of an open wagon full of coal.
  - An investigator reaches into a cage full of rabbits and selcts the first 10 that her hands touch, without conscious planning.
  - An investigator enlists the help of his Psychology 101 class to serve as "typical students" in a survey of average number of study hours worked by university students.

- For the following survey situation, *and in the context of the situation described*, carefully state the target population, the frame, and the sampling units. Also discuss any possible sources of selection bias or inaccuracy of responses, if appropriate.

  - A sample of 8 architects was chosen in a city with 24 architects and architectural firms. To select a survey sample, each architect was contacted by telephone in order of appearance in the telephone directory. The first 8 agreeing to be interviewed formed the sample.

- From the URSU Exam Database find the Stat 257 final exams from previous semesters. Solve as many of them as you can. Note, however, that not all of the problems from previous semesters are applicable to this semester.