Statistics 257.001 Fall 2004
December 6, 2004

1. 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.

2. Consider the following scenarios. Briefly exlain 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.)

- 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.

3. 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.

4. A University of Regina engineering professor is interested in building a model to study automobile fuel consumption. She begins by looking at the relationship between idle fuel consumption and engine capacity. The data she collected are given below:

idle fuel consumption (mL/s)	engine size (L)
0.17	1.2
0.32	1.8
0.38	2.5
0.51	3.4
0.62	4.1

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.