University of Waterloo STAT 332 – W. H. Cherry

Assignment 2

A2 – 1. Using equiprobable selecting, 2,070 farms were obtained from an unstratified study population of $\mathbf{N} = 20{,}700$ farms, and information was collected on the number of cattle ($\mathbf{Y}$) on *all* of the $n = 2{,}070$ farms; numerical summaries of the data obtained are given above.

$$\sum_{j=1}^{n} y_j = 25{,}881 \quad , \quad \sum_{j=1}^{n} y_j^2 = 599{,}486$$

(a) Estimate the average number of cattle per farm and the total number of cattle in the population; find also the estimated standard deviations (or estimated 'standard errors') of both estimates.

(b) Find an approximate 90% confidence interval for the average and for the total in (a).

(c) Comment briefly on the appropriateness of using equiprobable selecting in such a sample survey.

A2 – 2. In an audit of industrial accounts, 20 accounts were obtained by equiprobsable selecting from a total of 500 outstanding (*i.e.*, owing for more than 30 days) accounts; the average for all 20 accounts was \$105.00 and the standard deviation was \$47.33. Find an approximate 95% confidence interval for the total amount owing on overdue accounts.

A2 – 3. Using equiprobable selecting, two non-overlapping groups of size 200 and 450 were obtained successively without replacement from a study population of 2,400 students in a non-residential college. Each student in the groups was asked the distance (in km) that he or she lived from the college; the respective averages and standard deviations (based on complete response) were as shown at the right above. Find an approximate 99% confidence interval for the average distance from the college that its students live.

$$\bar{y}_1 = 5.14 \quad , \quad s_1 = 1.90$$
$$\bar{y}_2 = 4.90 \quad , \quad s_2 = 2.20$$

A2 – 4. An investigation was undertaken to compare the average salaries of females and males in an organization which employs 630 females and 720 males. Using equiprobable selecting, 70 females were obtained and showed an average salary of \$54,500 with a standard deviation of \$3,400; the corresponding values for 90 males were \$58,100 and \$3,900, all values being based on complete response.

(a) Assuming that the two selecting processes took place independently, find an approximate 90% confidence interval for the *difference* between the average salaries of females and males in the organization.

(b) On the basis of the interval found in (a), explain briefly whether these data provide evidence of a real (*i.e.*, non-zero) difference between the average salaries of females and males in this organization.

A2 – 5. A forester wishes to estimate the total value of the lumber in a stand of 5,000 trees. From a valuation of 50 trees obtained by equiprobable selecting, she finds an average of \$64 and a standard deviation of \$20. Approximately how many *further* trees should be selected by equiprobable selecting to be able to estimate the total value of the stand to within \$10,000 with approximately 95% probability?

A2 – 6. (a) Show that the function $f(p) = p(1-p)$ is a *maximum* when $p = ½$.

(b) A sample survey is conducted to estimate the proportion, $\mathbf{P}$, of people with a certain characteristic. The survey is carried out by obtaining, by equiprobable selecting, 1,024 people from a large study population; $\mathbf{P}$ is estimated as $p \equiv y/1{,}024$, where the random variable Y represents the number of people selected with the characteristic under investigation. Assuming no non-response, show that the approximate probability p is within 4 percentage points of $\mathbf{P}$ is at least 0.98, no matter what value $\mathbf{P}$ has.

A2 – 7. (a) A company auditor needed to estimate the proportion of travel vouchers that were wrongly filed. Of 50 vouchers obtained by equiprobable selecting from a total of 250 vouchers in a particular department of the company, 20 were found to be filed wrongly. Find an approximate 99% confidence interval for the proportion of wrongly filed travel vouchers in the department.

(b) For a residential area containing 5,000 private homes, it is desired to estimate the proportion of houses with (i) more than three people living in them; (ii) more than one car owned by the occupants. The estimators are required to have *standard deviations* not exceeding 0.02 and 0.01 respectively. From previous surveys, it is known that the proportions for (i) and (ii) lie in the respective ranges 0.35-0.55 and 0.10-0.20. If the two proportions are to be estimated in the *same* survey based on equiprobable selecting, find approximately how large a sample size is needed, assuming that there will be complete response.

A2 – 8. To determine the total volume of marketable lumber on a 1500-acre plantation, an investigator needs to estimate the total number of "count trees" (*i.e.*, trees larger than a specified size) on the plantation. Using an aerial photograph, the plantation is divided into 1,500 one-acre plots, and 100 of the plots are obtained by equiprobable selecting; from an actual inspection of these 100 plots, an average per plot of 25.2 count trees is found, with a sample standard deviation of 11.7 such trees. Give an approximate 95% confidence interval for the *total* number of count trees on the plantation.

(*continued overleaf*)

A2 – 9. Under the title of *Jobs outlook gloomiest since recession*, the *Toronto Star* of Monday, March 5, 1990, on page B1 reported the results of a sample survey of nearly 1,500 companies from across Canada by Manpower Temporary Services. The article presented a number of point estimates for proportions of respondents in various categories relevant to the matters under investigation, but no confidence intervals were provided. In an assessment of the survey, a student made the following comments:

(a) *Unfortunately, the article does not contain any confidence intervals ..... that indicate how likely it is the estimates about the job market are correct*;

(b) *It is hard to tell, without confidence intervals, whether the survey organization took into account that their results are based on a volunteer sample, as the article would seem to indicate.*

Critically discuss these two statements.

A2 – 10. A machine shop has 98 operators and the shop supervisor wishes to estimate the average time taken to perform a particular machining task. The supervisor obtains 8 operators by equiprobable selecting and times each of them as they do the task; the times (in minutes) are:

4.2 5.1 7.9 3.8 5.3 4.6 5.1 4.1.

(a) Discuss critically the method used by the supervisor to obtain the eight observations.

(b) Explain briefly whether the eight observations provide a proper basis on which to find an approximate 90% confidence interval for the average time needed to do the machining task in the machine shop.

A2 – 11. (a) A psychologist wished to estimate the average reaction time to a stimulus among 200 patients in a hospital specializing in the treatment of disorders of the nervous system. She obtained 20 patients by equiprobable selecting and measured their reaction times; the average was 2.1 seconds and the standard deviation was 0.20 seconds, based on complete response. Explain briefly whether these data provide a proper basis on which to find an approximate 95% confidence interval for the average reaction time for the 200 patients.

(b) Referring to (a) above, if it was desired to estimate the average reaction time of the 200 patients to the nearest one-tenth of a second with approximately 99% probability, find the (approximate) sample size that would be needed.