

Last name:

First name:

Student #:

**UNIVERSITY OF TORONTO
Faculty of Arts and Science**

APRIL/MAY 2008 EXAMINATIONS

STA 322H1 S

Duration - 3 hours

PLEASE HAND IN

Examination Aids: Non-Programmable Calculator, aid sheet, both sides, with theoretical formulas only.

[14] 1) The provincial government offers you to carry out a survey to investigate the financial problems of students at the University of Toronto. You have studied the problem in the preparatory phase and you have to clarify some issues. Try to give realistic but short answers to the following problems.

- (a) Describe the target population in some details. Should some types of students be excluded, with special types of financing or problems, or from some other reason, or they should be treated separately?
- (b) What variables may be of interest to the government in the study?
- (c) What parameters may be of interest? (**continued**)

- (d) (i) Would a simple random sample be a convenient design? Explain why, or why not. (ii) Suggest another sampling design you think is more appropriate for this study, and give some details on it regarding the population in this question. (iii) Should the three campuses of U of T be treated separately, or not? Would this improve or simplify your study in any sense?
- (e) List some methods of data collection and propose one you think might be appropriate here. Describe briefly some problems you may encounter in your method of data collection

[16] 2) From a small suburban community of 24 households, 5 were selected at random, and the number of adults, number of children, and total daily income were recorded for each.

Household	1	2	3	4	5
Number of adults, y	3	4	3	5	6
Number of children, z	2	4	1	4	4
Total daily income, x (in \$10)	33	40	34	68	61

($\sum y=21$, $\sum y^2=95$, $\sum x=236$, $\sum x^2=12190$, $\sum xy=1067$, $\sum z=15$, $\sum z^2=53$, $\sum zy=69$)

- (a) Estimate (i) the total number of residents in the community, (ii) total number of children, (iii) average household size, and (iv) total daily income in the community.
- (b) Estimate (i) the average monthly income per household and per adult (one month = 30 days), (ii) average number of children per household, (iii) percentage of children in the community and (iv) average percentage of children per household. **(continued)**

- (c) What kinds of estimators are used in (b) (answer for each case)?
- (d) (i) Place an error bound on the estimator of the average monthly income per adult in (b)(i). (ii) Place an error bound on the estimator in (b)(iv).

[20] 3) A market research firm conducted a survey in 1996 in a city for the purpose of estimating the total monthly household expenditures on compact discs (CDs) and the total number of households owning a compact disc player (CDP). The city was divided into four geographical areas and a random sample of households was selected from each area. The results of the survey are as follows:

Area	Number of Households	Number Sampled	Sample Average Monthly Expenditure (\$)	Sample Proportion Owning a CDP
Northeast	20,000	100	10.40	15%
Northwest	10,000	100	6.10	8%
Southwest	35,000	100	4.05	4%
Southeast	15,000	100	8.24	7%
Total	80,000	400		

- Estimate the average monthly household expenditure on CDs in the city, and the proportion of households in the city owning a CDP.
- Estimate the total monthly household expenditure on CDs in the city, and the total number of households owning a CDP in the city.
- How many households would be sampled from each area if the sample of 400 were with proportional allocation? Considering already obtained sample, do you think that the stratified sample with proportional allocation would produce better results than an SRS of the same size? (**continued**)

(d) On which parameter estimated in (a) you can place a bound on the error of estimation? Explain and calculate.

(e) What would be a bound on the error of estimation of the proportion of households in the city that own a CDP if you used the optimal (Neyman) allocation and the same sample size of 400? (use information from the sample)

[14] 4) An auditor is confronted with a long list of accounts receivable for a firm. She must verify the amounts on 10% of these accounts and estimate the average difference between the audited and book values. In the following cases explain your choice.

(a) Suppose the accounts are arranged chronologically, with the older accounts tending to have smaller values. Would you choose a systematic or a simple random sampling design to select the sample?

(b) Suppose the accounts are arranged alphabetically. Would you choose a systematic or a simple random sampling design to select the sample?

(c) Suppose the accounts are grouped by department and then listed chronologically within departments. Would you choose a systematic or a simple random sampling design to select the sample? **(continued)**

- (d) How would you estimate the variance of the sample mean according to your choice of the design in (a)?
- (e) How would you estimate the variance of the sample mean according to your choice of the design in (b)?
- (f) How would you improve the design in (a) to be able to safely estimate the variance of the sample mean.

[18] 5) A survey is conducted in a town to estimate the number of residents who visited physician and their expenditures on drugs in the month immediately preceding the survey. The survey made use of the available list of all residential units (the number of units being 643). Five of residential units were selected from the list at random, and all the residents of the unit were interviewed. The results are the following:

Residence No.	Number of residents	Number who visited a physician	Expenditures on drags in residence (\$)
45	3	0	0, 0, 65
131	2	1	400, 350
207	1	1	105
398	2	0	0, 0
519	2	1	50, 0

- (a) What is the population of interest in this study? What are the sampling units, what is the frame? What kind of sampling design was used here?
- (b) Estimate the number of residents in the town. Estimate the proportion and the number of residents in the town who visited a physician in the preceding month. Which of these estimators are biases, unbiased? Why?
- (c) Estimate the average expenditure on drugs per residence and per person and the total expenditure on drugs in the town in the preceding month. **(continued)**

- (d) Place a bound on the error of estimation of the total expenditure on drugs in (c).
- (e) Later was found that the town population was 1248. Use this information to reestimate the total expenditure on drugs in (c). Which of these two estimators would you prefer in general? In this particular case (don't calculate anything; take also into account the sample size)?

[18] 6) A psychologist wants to determine how many hours per week 8-10 year old boys in Toronto spend playing video and computer games. To investigate this she randomly selects three Toronto primary schools from the population of 230 primary schools in Toronto, and then randomly selects 20 (twenty) 8-10 year old boys from each school. With the parents help, she has each of the sampled boys record their video and computer game times for a calendar week. The results are presented in the following table:

School	Number of 8-10 year old boys	Sample Mean	Sample St. Dev.
A	350	15	5
B	200	12	4
C	150	11	3

- (a) Explain what kind of design is used here. Do you think this design is appropriate? Explain (ignore the small sample size problem).
- (b) Comment on the condition of the design "... *for a calendar week* ...". Would the selection of a "calendar week" affect results of the study?
- (c) Estimate the total number of hours spent by 8-10 year old Toronto boys during a week on video and computer games, and the variance of the estimator (you do not need to complete the calculation of the variance, but you have to calculate all entries). Is this estimator unbiased? Explain. **(continued)**

- (d) Do you expect that your standard error in (c) would be larger, smaller or about the same as the standard error of an estimator based on SRS of the same size? Explain.
- (e) Estimate the percentage of 8-10 year old boys free time spent during a week on playing the games. Assume that 8h a day is spent on sleep (7 days a week), and 7h in school and transportation (5 days a week).

Total marks = 100