Last name: First name: Student #:

STA 304H1 F/1003H F SUMMER 2009, First Test, May 28 (20%)
Duration: 50min. Allowed: hand-calculator, aid-sheet, one side, with theoretical formulas and definitions only.

[25] 1) A marketing analyst is asked to study the buying habits of shoppers at a national chain store (e.g. Sears). Suppose there are 150 stores around the country.
(a) Describe the population of interest.
(b) Describe in short **a realistic** sampling procedure for obtaining a representative sample in this problem, and give a name of the procedure.
(c) Are the target population and sampled population the same? Explain some related problems.
(d) Give two variable of interest related to element of the population (one quantitative, the other qualitative).
(e) Describe an appropriate method of data collection in this study.

Solutions:
[5](a) All shoppers at the chain store. More accurate definition would be: All shoppers that regularly shop at the chain store, but then it should require to defining who is a "regular shopper". The definition may also include a time period of shopping. **[5]**

[7](b) Two stage cluster sampling: First select an SRS of stores, and then a sample of customers, e.g. when entering the store, or at exit, using systematic sampling, because a list of shoppers does not exist. **[7]** Selecting customers from each store is possible but would be inconvenient and much more costly. Also, the sampling design may include a rule of selecting a sampling day at the store. Sampling at the same day at every store may produce a bias result. It is also possible to send a questionnaire to the registered customers (most of them may have the chain's membership card), but this would also be very costly and with a large nonresponse rate.

[6](c) They are not the same. If the sample is taken from the shoppers present at stores, that the sampling population is much smaller that the target population. If the sample is taken from registered shoppers, many shoppers may not be registered. **[6]**

[4](d) Quantitative: amount of money spent in one visit/a year, number of items purchased, … Qualitative: item of primary interest, items of interest found, satisfaction, … **[4]**
[3] (e) A short personal interview, or short questionnaire filled in at the spot. **[3]**

[25] 2) A list contains 600 names of 300 married couples that donated to a local charity. Names are listed in alphabetic order and assigned numbers 1,2, ... ,600, couple by couple, first husband and then wife. Use the following portion of the table of random numbers
23632 27889 47914 02584 37680 20801 72152 39339 34086 43218 15263 31624 76384
17403 53363 44167 64486 64758 75366 76554 31601 12614 33072 60332 92325 19474
01624 76384 97403 53363 44167 64486 64758 75366 76554 31601 12614 33072 19474
23632 27889 47914 02584 37680 20801 72152 39339 34806 08930 25570 33120 45732
in the following problems:
(use some simple procedure and explain it briefly in each of the cases):

(a) Select an SRS of 5 **couples** from the list (give the list numbers of these 10 persons)
(b) How would you select an SRS of 5 **wives** from the list?
(c) How would you select an SRS of 5 **persons** (regardless of the gender) from the list?
(d) If you want to estimate the average life of the people in this group, could it be appropriate to use any of the samples from (a), (b), or (c)? Explain. How about estimating the average length of marriage?

Solutions:
[7] (a) Two numbers from the list are assigned to each couple, i.e. we have 300 pairs. So we need to select 5 pairs out of 300. We may assign random numbers 001, 002 to pair 1, 003, 004 to par 2, …, 599, 600 to pair 300 **[3]**:
Using groups of three digits and, say, the third row, we have from the table

digits	016	247	638	497	403	533
couple	015	247	N/A	497	403	533
	016	248	N/A	498	404	534

Sample (list numbers): 15, 16, 247, 248, 403, 404, 497, 498, 533, 534. **[4]**

[6] (b) You can select 5 wives in the same way as you have selected 5 couples, and list only wives' numbers. Or, you can use only three digits ending by even digit, because all wives are listed with even numbers. **[6]** E.g., from the first row, we have digits
236, **322**, 788, 947, 914, 625, 843, 768, **376,** 802, **080**,**172**. Using only "even" numbers less than equal to 600, we obtain the sample of 5 wives: 80, 172, 236, 322, 376 (sample not required).

[6] (c) In this case, the population is of size 600, and we just can use three digits, up to 600, e.g. from the third row: 174, 035, 336, 344, 167. Sample numbers: 35, 167, 173, 336, 344. **[6]** (sample not required)

[6] (d) The sample from (c) is an SRS from people and then is appropriate to use it to estimate the average life. In (a) and (b) the information about the age of the couple can be also obtained, but the values of couple's ages are highly correlated, and the sample would not be an SRS, and the result might be biased.**[3]** For the length of the marriage, the information is related to the couples, and then both the samples from (a) and (b) are appropriate (SRS of couples, or wives), but in (c) the people from the same couple may appear in the sample and then make sample biased.**[3]**

[30] 3) There are 12 grocery stores in a city block with numbers of employees as follows:

store	1	2	3	4	5	6	7	8	9	10	11	12
# empl.	3	5	4	4	2	3	6	4	5	3	7	5

(a) Explain in short how you would use the table of random numbers to select an SRS of size 4 (without replacement) from this population.
(b) An SRS of size 4 was selected and stores 2, 4, 6, 10 were obtained. Estimate the population mean and the population variance from the sample using unbiased estimators.
(c) Place a bound on the error of estimation of the total number of employees.
(d) Calculate the exact error of your estimation of the total and the theoretical variance of $\hat{\tau}$.

Solutions:
[6](a) N = 12, and then we have to use two digits from the table of random numbers, e.g. consecutive pairs, and select store 1, if 01 appears, store 2, if 02 appears, …, store 12, if 12 appears. We have to ignore all other pairs, 13,14, …, 99, 00. **[4]** If we want to be more efficient in reading the table, we may assign 5 pairs of digit to each store, e.g., using

Select store	1	2	…	12
Random digits	01	02		12
	21	22		32
	41	42		52
	61	62		72
	81	82		92

We still have to ignore 13, 14, …, 20, 33, 34, …, 93, 94, …, 99, 00. **[2]** Other similar methods may be used.

[9](b) Sample values: 5, 4, 3, 3. $\hat{\mu} = \bar{y} = 3.75,$ **[4]**

$S^2 = 0.917, \hat{\sigma}^2 = \frac{N-1}{N} S^2 = \frac{11}{12} 0.917 = 0.840$ (unbiased). **[5]** (If the answer is only 0.917, give **3**)

[6](c) $B_\tau = 2N\sqrt{\frac{N-n}{N}\frac{S^2}{n}} = 2\times 12 \times \sqrt{\frac{12-4}{12}\frac{0.917}{4}} = 18.77$. **[6]**

[9](d) $\tau = 51$ (from the list), $\hat{\tau} = 12\hat{\mu} = 12 \times 3.75 = 45,$ the error = $|\tau - \hat{\tau}| = 6$. **[4]**

For $Var(\hat{\tau}) = N \times \frac{N-n}{N-1}\frac{\sigma^2}{n}$ calculate first σ^2 from the population.

$\sigma^2 = 1.854$

$Var(\hat{\tau}) = 12 \times \frac{12-4}{12-1} \times \frac{1.854}{4} = 12 \times 0.337 = 4.044$. **[5]**

[20] 4) Suppose you wanted to estimate the proportion of adults who write with their left hands and decide to watch a reasonably selected sample of *n* people signing credit card receipts at a big store (e.g., Sears from Q.1).
(a) Do you think this sample would be representative for the question of interest? Discuss.
(b) Assuming the sample is representative, how many people would you have to include in the sample to estimate the proportion with a margin of error of 5%. Your guess is that the proportion cannot be greater than 30%.
(c) Suppose you observed 500 people and 100 of them signed with their left hand. What interval of values would you be 95% confident that contain the true proportion of left-handed people in the adult population.

Solutions:
[7](a) Assuming that left-handed people sign with their left hand (which is not for certain), and that being a left-handed is randomly distributed over the population (at least we don't expect that having a credit card and shopping in a big store is related to be a left-handed), we may expect the sample be representative of the population. **[7]**

[7](b) $D = (0.05/2)^2 = 0.000625$, N is large, so that
$n = \sigma^2 / D = pq / D = 0.3 \times 0.7 / D = 336$, where $p \le 0.3$ is used. **[7]**
A more conservative value would be obtained if no information about p is assumed, that is if p = 0.5 is used, and then n = 400.

[6](c) 95% CI for p is $\hat{p} \pm 2\hat{\sigma}(\hat{p}) = 100/500 \pm 2\sqrt{\hat{p}\hat{q}/500} = 0.20 \pm 0.0358$
$\hat{p} \pm 2\hat{\sigma}(\hat{p}) = 60/500 \pm 2\sqrt{\hat{p}\hat{q}/500} = 0.12 \pm 0.03$ =[0.09;0.15]. **[6]**