STATISTICS 4310/7310
SAMPLE EXAM I

NAME: _____

ID NUMBER: _____

Problem	Points Possible	Points Received
1	30	
2	30	
3	40	
Total	100	

INSTRUCTIONS:

1. Fill in your name, ID number in the spaces provided above.
2. The Exam is closed book and closed notes, except that you are allowed to have 3 sheets of notes you prepared.
3. Carry all computation to at least two decimal places.
4. The points for each question are indicated in the left margin in square brackets [ ].
5. **BE SURE TO SHOW ALL RELEVANT FORMULAS AND WORK!!**

1. An SRS of size 30 is taken from a population of size 100. The sample values are: $8, 5, 2, 6, 6, 3, 8, 6, 10, 7, 15, 9, 15, 3, 5, 6, 7, 10, 14, 3, 4, 17, 10, 6, 14, 12, 7, 8, 12, 9$

[10] (a) What is the sampling weight for each unit in the sample?

Answer: 10/3

[10] (b) Use the sampling weights to estimate the population total, t.

Answer: $\hat{t} = \sum_{i=1}^{3} 0(10/3)y_i = (10/3) * 247 = 823.33$

[10] (c) Given a 95% CI for t. Does the fpc make a difference for this sample?

Answer: $s^2 = 15.97816, \hat{V}(\hat{t}) = 3728.238, SE(\hat{t}) = 61.0593, \Rightarrow 95\%CI[703.65, 943.01]$
fpc $= 1 - n/N = 0.7$*, so it reduces the width of the CI.*

[30] 2. For the following survey, describe the target population, sampling frame, sampled population, sampling unit, and observation unit. Discuss any possible sources of selection bias.

The American Statistical Association sent the following e-mail with subject line "Joint Statistical Meetings 2005 Participants Survey" to a sample of persons who attended the 2005 Joint Statistical Meetings:"Thank you for attending the 2005 Joint Statistical Meetings (JSM) in Minneapolis, Minnesota. We need your help to complete an online survey about the JSM. Because the quality of the JSM is very important, a survey is being conducted to find out how we might improve future meetings. We would like to get your opinion about various aspects of the 2005 meeting your preferences for 2006 and beyond.
You are part of a small sample of conference registrants who have been selected randomly to participant in the survey. We hope you will take the time to complete this short questionnaire online at www.amstat.org/meetings/jsm/2005/survey. In order to tabulate and analyze the data, please submit your response by mid-September 2005."

Answer:
Target population: all the attendees of 2005 JSM.
Sampling frame: email addresses provided by the attendees of the 2005 JSM.
Sampled population: the attendees who provided their email addresses.
Sampling unit: One e-mail address.
Observation unit: person

It is stated that the small sample of conference registrants was selected randomly. This is good, since the ASA can control the quality better and follow up on non- respondents. It also means, since the sample is selected, that persons with strong opinions cannot flood the survey. But nonresponse is a potential problem—response is not mandatory and it might be feared that only attendees with strong opinions or a strong sense of loyalty to the ASA will respond to the survey.

3. Hayes (2000) took a stratified sample of New York City food stores. The sampling frame

consisted of 1408 food stores with at least 4000 square feet of retail space. the population of stores was stratified into three strata using median household income within the zip code. The prices of a "market basket" of goods were determined for each store; the goal of the survey was to investigate whether prices differ among the three strata. Hayes used the logarithm of the total price for the basket as the response y. Results are given in the following table.

Stratum, h	N_h	n_h	$\bar{y}_h$	s_h
1 Low income	190	21	3.925	0.037
2 Middle income	407	14	3.938	0.052
3 Upper income	811	22	3.942	0.070

[10] (a) The planned sample size was 30 in each stratum; this was not achieved because some stores went out of business while the data were being collected. What are the advantages and disadvantages of sampling the same number of stores in each stratum?

Answer: An advantage is that using the same number of stores in each stratum gives the best precision for comparing strata if the within-stratum variances are the same. In addition, people may perceive that allocation as fair. A disadvantage is that estimates may lose precision relative to optimal allocation if some strata have higher variances than others.

[15] (b) Estimate $\bar{y}_U$ for these data and give a 95% CI.

Answer: $\bar{y}_{str} = 3.9386, and 95\% CI is [3.92; 3.96]$

[15] (c) How would you allocate the sample sizes among three strata? Why?

Answer: optimal allocation since the stratum variance is different.