

**BST 5420**  
**Homework 3**  
**Due April 2, 2019**

This is worth 51 points, and the recorded score will be the proportion correct times 10 (rounded to the nearest half point) in order to be on a 10 point scale.

**1. (12 Points)** A researcher developed a test designed to measure the degree of awareness of current public health events. She wants to estimate the average score that would be achieved on this test by all students in a certain high school. The administration at the school will not allow the experimenter to randomly select students from classes in session, but it will allow her to interrupt a small number of classes for the purpose of giving the test to every member of the class. Thus, the experimenter selects 25 classes at random from the 108 classes in session at a particular hour. The test is given to each member of the sampled classes, with results as shown in the table below. Estimate the mean score that would be achieved on the test by all students in the school. Also, give an approximate 95% confidence interval for the mean.

Class	Number of Students	Total of all Scores	Class	Number of Students	Total of all Scores
1	31	1590	14	40	1980
2	29	1510	15	38	1990
3	25	1490	16	28	1420
4	35	1610	17	17	900
5	15	800	18	22	1080
6	31	1720	19	41	2010
7	22	1310	20	32	1740
8	27	1427	21	35	1750
9	25	1290	22	19	890
10	19	860	23	29	1470
11	30	1620	24	18	910
12	18	710	25	31	1740
13	21	1140			

**2. (21 Points)** Read the paper

Thorpe, Lorna E., et al. "Childhood obesity in New York City elementary school students." *American Journal of Public Health* 94.9 (2004): 1496-1500.

This can be obtained by searching "thorpe childhood obesity new york" at [scholar.google.com](https://scholar.google.com).

The authors say "This height-and-weight survey was conducted with a stratified, multistage, probability sample of elementary public school children in New York City."

(a) Describe what "stratification" means in this context. What are the strata?

(b) Describe what multistage means in this context. Within each stratum, what is the design of the sampling plan?

- (c) Describe the levels of the sampling design from part (b). (For example, “four-stage cluster sample where a sample of \_\_\_\_ psus is selected from \_\_\_\_; then from each \_\_\_\_ a sample of \_\_\_\_ from ... ”)
- (d) Describe what probability sample means in this context.
- (e) Describe one of the response variables in this study.
- (f) What is the population in this study?
- (g) Describe one parameter that is estimated (e.g., mean, total, proportion). Give its point estimate and margin of error.

3. **(12 Points)** Determine the sampling weights for the following sampling plans.

- (a) A simple random sample of size  $n$  from a population of size  $N$ .
- (b) A stratified random sample where for stratum  $h$  ( $h = 1, 2, \dots, H$ ),  $n_h$  units are selected from the  $N_h$  items in.
- (c) A one-stage cluster sample with notation given on p. 169, where each cluster is selected with equal probability.
- (d) A one-stage cluster sample where the probability of selecting a cluster is proportional to the cluster size, and clusters are selected with replacement.

4. **(6 Points)** Read problem 2 on pp. 267-268.

- (a) Don't work out the problem as stated in the book, but rather, use R to select a sample of size 10 with replacement using the probabilities  $\psi_i$ . Show your R code.
- (b) Ignoring the probabilities  $\psi_i$  use R to select a sample of size 10 with replacement. Show your R code.