Chapter 11

1) a) \( \bar{Y}_{st} = \frac{100}{450} \cdot 10 + \frac{50}{450} \cdot 20 + \frac{300}{450} \cdot 30 = 24.44 \)

b) 95% CI: \( \bar{Y}_{st} \pm z \cdot \sqrt{\text{var}(\bar{Y}_{st})} \) as all \( n > 30 \)

\( \text{var}(\bar{Y}_{st}) = \sum_{h=1}^{k} \left( \frac{N_h}{N} \right)^2 \cdot \left( 1 - \frac{n_h}{N_h} \right) \cdot \frac{S_h^2}{n_h} = 5.83 \)

\( 24.4 \pm \sqrt{5.83} = (19.713, 29.176) \)

2) a) Proportional Allocation

\( n_1 = \frac{N_1}{N} \cdot n = \frac{200}{500} \cdot 100 = 40 \)

\( n_2 = \frac{N_2}{N} \cdot n = \frac{300}{500} \cdot 100 = 60 \)

b) Optimal Allocation

\( n_1 = \frac{\frac{N_1 \sigma_1}{\sum_{i=1}^{k} N_i \sigma_i}}{n} = \frac{200 \cdot 9}{3000} \cdot 100 = 40 \)

\( n_2 = \frac{\frac{N_2 \sigma_2}{\sum_{i=1}^{k} N_i \sigma_i}}{n} = \frac{300 \cdot 4}{3000} \cdot 100 = 60 \)

3) a) Some problems include labeling the units and insuring randomness in the collection of the data. Here I examined a 3 strata population consisting of 1st, 2nd, and 5th graders of a particular school district, and how far they can throw a standard medicine ball.

b) Here, I decided to simulate the data, with the assumption that the data from first graders \( \sim N(5,1^2) \), third graders \( \sim N(10,2^2) \), and fifth graders \( \sim N(15,4^2) \). I decided to sample 30 first graders, 20 third graders, and 25 fifth graders. We also knew from school records, there are 150 first graders, 200 third graders, and 250 fifth graders in the school district.

c) If the data here were actually collected, accurate distance measurement and the willingness of the students will be a problem. Randomness would also be a problem here, if we were to simply collect the data from a few gym classes.

d) \( \bar{Y}_{st} = \frac{200}{600} \cdot 10.256 + \frac{150}{600} \cdot 4.942 + \frac{250}{600} \cdot 15.582 = 11.072 \)

e) \( \text{var}(\bar{Y}_{st}) = \sum_{h=1}^{k} \left( \frac{N_h}{N} \right)^2 \cdot \left( 1 - \frac{n_h}{N_h} \right) \cdot \frac{S_h^2}{n_h} = .10376 \)

f) \( 11.072 \pm \sqrt{.10376} = (10.760, 11.394) \)
g) Proportional Allocation
\[ n_1 = \frac{N_1}{N} \cdot n = \frac{150}{600} \cdot 200 = 50 \]
\[ n_2 = \frac{N_2}{N} \cdot n = \frac{200}{600} \cdot 200 = 67 \]
\[ n_3 = \frac{N_3}{N} \cdot n = \frac{250}{600} \cdot 200 = 83 \]
Optimal Allocation
\[ n_1 = \left( \frac{N_1 s_1}{\sum N_i s_i} \right) \cdot n = \frac{130.62}{1456.02} \cdot 200 = 18 \]
\[ n_2 = \left( \frac{N_2 s_2}{\sum N_i s_i} \right) \cdot n = \frac{1141.4}{1456.02} \cdot 200 = 61 \]
\[ n_3 = \left( \frac{N_3 s_3}{\sum N_i s_i} \right) \cdot n = \frac{884}{1456.02} \cdot 200 = 122 \]

h) Take a larger sample to help minimize standard error, and to improve randomness, if these numbers were taking a cluster sample of a single gym class

Chapter 12

1) a) \[ \hat{\tau} = \frac{N}{n} \sum_{i=1}^{n} y_i = \frac{10}{35} \cdot (4 + 12 + 7) = 76.67 \]
   b) \[ \text{var}(\hat{\tau}) = N \cdot (N - n) \cdot \frac{s^2}{n} = 381.111 \]

2) a) \[ \hat{\tau}_r = M \cdot \frac{\sum y_i}{\sum M_i} = 100 \cdot \frac{23}{35} = 65.714 \]
   b) \[ \text{var}(\hat{\tau}_r) = \frac{N(N - n)}{n(n - 1)} \cdot \sum_{i=1}^{n} (y_i - rM_i) = \frac{70}{6} \cdot 2 = 23.33 \]

3) a) \[ \hat{\tau}_r = \frac{M}{n} \cdot \sum \frac{y_i}{M_i} = \frac{100}{3} \cdot \frac{21}{10} = 70 \]
   b) \[ \text{var}(\hat{\tau}_r) = \frac{M^2}{n(n - 1)} \cdot \sum \frac{(y_i - \hat{\mu}_r)^2}{\hat{\mu}_r^2} = \frac{5000}{3} \cdot \frac{1}{50} = 33.33 \]
5) a) \[ \hat{\mu} = \frac{10}{1} (1 + 0 + 2 + 3 + 0 + 1) = 70 \]

This design is a systematic approach, hence the appropriate estimator.

b) In the question, they used a ratio estimator which is not unbiased with this sampling scheme.

c) It is not unbiased with a sample size of 1 with schematic sampling, as with only a single unit sample size, it is impossible to have an unbiased estimator.

7) In the realm of equally likely draws of primary units in the context of this problem,

\[ \gamma = \text{sum of all secondary units} > \text{sum of all primary units} \]

Proposed estimator: \[ \hat{\mu} = N \cdot \bar{Y}(i) \]

\[ E(N \cdot \bar{Y}(i)) = N \cdot \frac{1}{N} \sum_{i=1}^{N} \bar{Y} = \gamma \]

\[ \text{Var}(\hat{\mu}) \text{ exists, but cannot be found with a sample size of 1.} \]

In proposed model, \( \gamma \) scores are i.d.

\[ \Rightarrow \text{Each list of sample same as \( N \) can be viewed as selecting a \( \text{SRS} - 1 \text{m from} \text{M} \) \}

A giving \( P(1) \) their scores.