SURVMETH / SURV 625
Applied Sampling
Winter/Spring 2025

Homework 1

Download the Excel file "fem1524_admin.xlsx" from the homework folder on the course Canvas site. This file is a population list of $N = 2{,}920$ young women between the ages of 15 and 24, which will be considered as a sampling frame for this first homework assignment.

1. Select a simple random sample (SRS) of size $n = 20$ from this frame. **Each student will select a different simple random sample**, using the R code *set.seed(the last four digits of your UM/UMD student ID).* Note that we are simulating the notion of hypothetical repeated random sampling using the same SRS design! The class has 30 enrolled students and would generate 30 samples.

2. Give, in selection order, the list of the 20 four-digit selection number (IDs) and the values of AGE for the women in your sample.

3. Compute the sample estimate of the mean age. What else would we need to compute (<u>be specific</u>) to make inference about the mean age of the population?

4. What would we call the distribution that we would see if we plotted all 30 sample estimates of the mean age (computed from the 30 unique samples generated by the students in the class)? What would we call the standard deviation of this distribution?

5. Based on the ID numbers of the SRS sample that you selected above, use the data file available for this homework "SM 625 HW 1.xlsx" and work on the following questions.

   a) Look up the number of male sexual partners in the past year (PARTS1YR) that were reported in a survey by each of your 20 selections in the Excel file. Estimate the mean number of partners in the past year for the population, $\bar{Y}$ $(\bar{y} = y/n = \sum_{i=1}^{20} y_i / n)$.

   b) Estimate the population element variance, S^2 $[s^2 = \sum_{i=1}^{20} (y_i - \bar{y})^2 / (n-1) = \left(\sum_{i=1}^{20} y_i^2 - \frac{y^2}{n} \right) / (n-1)]$.

   c) Estimate the sampling variance of the mean, $Var(\bar{y})$, and the standard error, $SE(\bar{y})$ $\left[\text{var}(\bar{y}) = (1-f)\frac{s^2}{n} \text{ and } se(\bar{y}) = \sqrt{\text{var}(\bar{y})} \right]$.

   d) Compute a 95% confidence interval for the sample mean $(\bar{y} \pm t_{(1-\alpha/2, n-1)} \cdot se(\bar{y}))$.

   e) Explain why the mean computed in a) will generally not be equal to the population mean.

f) Estimate the coefficient of variation of the mean, $CV(\overline{y})$ $(cv(\overline{y}) = se(\overline{y}) / \overline{y})$.

g) What difference would it make for the sampling variance of the mean if the sample size were increased to $n = 60$?

h) What sample size is needed to obtain $se(\overline{y}) = 0.05$? What about a $cv(\overline{y}) = 0.10$? What about a 95% confidence interval with width 0.40 (using 2 for the t-value)?

i) Estimate the mean number of male sexual partners in the past year (and its standard error) for the subclass of teenagers (age 15-19) in the sample. Ignore the finite population correction in the calculation of the standard error. How does this standard error compare to the standard error for the full sample? Would you expect such a difference? If so, why?