

Case Study 1: National Youth Tobacco Study

The National Youth Tobacco Survey (NYTS) is a complex survey carried out by the CDC to better understand teen tobacco use patterns. It is a school-based, pencil-and-paper questionnaire, self-administered to a nationally representative sample of students in grades 6-12 in the U.S. In 2019, 5,675 students completed the NYTS; the overall survey response rate was 71.6%. Students were asked if they had used an e-cigarette in the past 30 days. In an effort to address the reasons students used e-cigarettes, they were also asked who introduced them to e-cigarettes and whether they used other forms of tobacco.

1. Describe the sample. How large is it (n) and what is the unit of observation?
2. Describe the population. Approximately how large is it (N)?
3. Select a population parameter that researchers would be interested in estimating and the corresponding statistic that will serve as the estimate.
4. Identify possible sources of statistical bias and whether they are selection, measurement, and non-response bias. How do you expect this will effect the estimate of the parameter?
5. Identify possible sources of variation and whether they are sampling or measurement variability. How do you expect this will effect the estimate of the parameter?
6. On a separate sheet of paper (a third page to your submission), make two sketches: 1. an empirical distribution of the data and 2. what you think the sampling distribution of the statistic might look like given what you know about the data collection. Be sure to label the axes of both plots. Add a vertical line on the sampling distribution that hits the x-axis where you expect the population parameter to be.

Case Study 2: World Health Organization

Substantial global progress has been made in reducing childhood mortality since 1990. The total number of under-5 deaths worldwide has declined from 12.6 million in 1990 to 5.0 million in 2020. Yet, two regions, Sub-Saharan Africa and Southern Asia, account for more than 80% of these under-five deaths, despite making up only 50% of the global population. These statistics are based on country data collected by the WHO. The WHO aims to use statistics from the 194 countries in the United Nations to associate child mortality with diseases such as infectious diseases, pre-term birth complications, and congenital anomalies. Each year the WHO collects a new data set and calculates new statistics. In the most recent year, the data came from 80 countries that provided birth registrations and data from their federal health departments.

1. Describe the sample. How large is it (n) and what is the unit of observation?
2. Describe the population. How large is it (N)?
3. Select a population parameter that researchers would be interested in estimating and the corresponding statistic that will serve as the estimate.
4. Identify possible sources of statistical bias and whether they are selection, measurement, and non-response bias. How do you expect this will effect the estimate of the parameter?
5. Identify possible sources of variation and whether they are sampling or measurement variability. How do you expect this will effect the estimate of the parameter?
6. Add to your third page two more sketches: 1. an empirical distribution of the data and 2. what you think the sampling distribution of the statistic might look like given what you know about the data collection. Be sure to label the axes of both plots. Add a vertical line on the sampling distribution that hits the x-axis where you expect the population parameter to be.

