

Lab Homework 5—Sampling Distributions

Youth Risk Behavior Surveillance System (YRBSS)

Every two years, the CDC conducts national surveys in schools to monitor and assess the six largest contributors to youth morbidity and mortality. These contributors include health risks, such as high body mass index, and risky behaviors, such as tobacco and alcohol use, drunk driving, and failure to use seat belts. In 2013, 47 states participated in this school-based survey titled the Youth Risk and Behavior Social Survey, yielding 13,583 respondents and 213 variables. A subset of this data set with no missing data for 16 selected variables is provided in the file `yrbss2013.csv`. You can access the complete survey and data documentation on the CDC website.

Variable	description
age	Q1: How old are you?
gender	Q2: What is your sex?
height_m	calculated variable: height in meters
weight_kg	calculated variable: weight in kilograms
bmi	calculated variable: body mass index=height m/(weight kg) ²
BMIPCT	calculated variable: BMI percentile for age and sex
seatbelt	Q9: How often do you wear a seat belt when riding in a car driven by someone else?
seatbelt2	calculated variable: seatbelt never vs otherwise
ride_drunkdriver	Q10: During the past 30 days, have you ridden in a car or other vehicle driven by someone who had been drinking alcohol?
drive_drunk	Q11: During the past 30 days, how many times did you drive a car or other vehicle when you had been drinking alcohol?
drive_text	Q12: During the past 30 days, on how many days did you text or e-mail while driving a car or other vehicle?
carried_weapon	Q13: During the past 30 days, did you carry a weapon such as a gun, knife, or club?
unsafe_school	Q16: During the past 30 days, did you not go to school because you felt you would be unsafe at school or on your way to or from school?
bullied	Q24: During the past 12 months, have you ever been bullied on school property?
sad	Q26: During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?
days_smoke	Q33: During the past 30 days, on how many days did you smoke cigarettes?
days_drink	Q43: During the past 30 days, on how many days did you have at least one drink of alcohol?

Lab Practice

Part One: Sampling Distribution for Proportions

1. Import the `yrbss2013.csv` dataset.
 - a. Pick any two categorical variables in the dataset (do not select the bullied variable) and give their frequency distributions. Report the proportions for both variables.
2. For each variable selected in question 1, create 4 different for loops:

- a. Create a for loop that gets 100 samples of size 10 from the population. Report the average proportions.
- b. Create a for loop that gets 5,000 samples of size 10 from the population. Report the average proportions.
- c. Create a for loop that gets 200 samples of size 20 from the population. Report the average proportions.
- d. Create a for loop that gets 200 samples of size 5000 from the population. Report the average proportions.
- e. Which one seems to have more of an effect on the sampling distribution: increasing the number of samples or increasing the sample size? Explain what that effect is.