Lab 8 Homework: Sampling Distribution of the Mean and Inference for a Single Mean

Part One: Sampling Distribution of the Mean

Data: yrbss2013.csv

Every two years the CDC conducts national surveys in schools to monitor and assess the six largest contributors to youth morbidity and mortality. These include not only health risks such as high BMI, but also risky behaviors such as tobacco and alcohol use, drunk driving, and texting while driving. In 2013, 47 states participated in this survey, yielding 13,583 respondents and 213 variables.

For this assignment, we focus on the variable `drive_text`, which records:

> *During the past 30 days, on how many days did you text or e-mail while driving a car or other vehicle?*

1. Consider the `drive_text` variable:

    (a) Load and summarize `drive_text`. Report the mean and standard deviation. Plot its distribution. What can you say about its shape?

2. Using `drive_text`, create four different for loops:

    (a) Draw 100 samples of size 10. Report the mean and standard deviation of the sample means. Check the distribution.

    (b) Draw 5,000 samples of size 10. Report the mean and standard deviation of the sample means. Check the distribution.

    (c) Draw 200 samples of size 20. Report the mean and standard deviation of the sample means. Check the distribution.

    (d) Draw 200 samples of size 5,000. Report the mean and standard deviation of the sample means. Check the distribution.

    (e) Which has more of an effect on the sampling distribution: increasing the number of samples or increasing the sample size? Explain.

Part Two: Inference for a Single Mean

Data: lead.csv

Research on the harmful effects of lead poisoning began in the 1970s. Lead poisoning affects the development of the nervous system and has a higher impact on children than adults. Exposure can be occupational or recreational through contaminated air, soil, water, or food. Effects include learning disabilities and behavioral problems. High levels can lead to seizures, coma, and death.

The data presented are from one of the first quantitative research articles on the topic, published in *The Lancet* in 1975. 124 children living near a lead-emitting smelter in El Paso, Texas were studied for two years. Tests included IQ tests and various neurological assessments.

Variables of Interest

- `Iqf`: Full scale IQ

- `Ld73`: Blood lead level in 1973 (micrograms/100mL)

1. Import the `lead.csv` dataset. Explore the mean IQ of children in the sample.

    (a) What is the mean IQ and standard deviation?

    (b) Is IQ normally distributed?

2. Another research study reports an average IQ of 85.

    (a) What test should you run?

    (b) State the null and alternative hypotheses.

    (c) Report the test statistic and p-value.

    (d) Report the confidence interval.

    (e) Do you reject or fail to reject the null hypothesis? Justify your answer.

3. Another study reports an average blood lead level in 1973 of 36.

    (a) What test should you run?

    (b) State the null and alternative hypotheses.

    (c) Report the test statistic and p-value.

    (d) Report the confidence interval.

    (e) Do you reject or fail to reject the null hypothesis? Justify your answer.