

Statistics 2025/4025 Homework 1

1. Suppose a researcher would like to determine whether one grade of gasoline produces better gas mileage than another grade. Twenty cars are randomly divided into two groups, with 10 cars receiving one grade and 10 receiving the other. After many trips, average mileage is computed for each car.

a. Would it be easier to detect a difference in gas mileage for the two grades if the 20 cars were all the same size, or would it be easier if they covered a wide range of sizes and weights? Explain.

b. Suppose thirty cars were used instead, randomly divided into one group with 20 cars receiving one grade and another group with 10 cars that received the other grade. Would that bias the results in any way? Explain.

2. Suppose researchers were interested in determining the relationship, if any, between brain cancer and the use of cellular telephones.

a. If either were possible, would it be better to use a randomized experiment or a case-control study? Explain.

b. Could you design your study so that it is double blind? Explain.

3. The data shown in the in the table below represent employees laid off by the U. S. Department of Labor. Make the assumption that these employees are representative of a larger population of employees.

Ethnic Group	Laid Off	Not Laid Off	Total
African American	130	1382	1512
White	87	2813	2900
Total	217	4195	4412

a. I calculated a 95% confidence interval for the true odds ratio and got (4.5,5.8). Explain what this means in terms of probability. (In other words, something is true with probability 0.95. What is that something?)

b. Is Simpson's Paradox a concern in this situation? Explain.

4. Suppose a researcher wanted to find out which of two chemicals works best to increase oak tree biomass. She conducted a controlled experiment involving 200 trees, randomly assigned half of them to each treatment (chemical #1 or chemical #2). Five days after applying the chemicals, she recorded the data below.

Chemical #1: $n=100$, mean biomass=23.6 standard deviation=5.8

Chemical #2: $n=100$, mean biomass=24.8 standard deviation=6.7

(a) The p-value from a t- test here is 0.18. What do you conclude? Explain.

(b) The 0.18 is a probability. What is the probability of?

(c) If the standard deviations had been 0.58 and 0.67 would the p-value be bigger or smaller? Explain.

5. Which of the following statements is not true regarding a 95% confidence interval for the mean of a population?

a. In 95% of all samples, the sample mean will fall within 2 standard errors of the true population mean.

b. 95% of the population values will lie within 2 standard errors of the sample mean.

c. In 95% of all samples, the true population mean will be within 2 standard errors of the sample mean.

d. If you add and subtract two standard errors to/from the sample mean, in 95% of all cases you will have captured the true population mean.

6. Which of the following conclusions do you draw if the p-value is smaller than 0.05?

a. The null hypothesis is false

b. The alternative hypothesis is true.

c. The test statistic is significantly different from the null value.

d. All of the above.