



## 2494 - COMPUTATIONAL THINKING & DATA SCIENCE

2019-20, Spring Semester

### In-class Exercises

#### SAMPLING

1. The file Accounts Receivable.xlsx contains 280 accounts receivable for the Spring Mills Company. There are three variables:

- **Size:** customer size, categorized as small, medium, or large depending on its volume of business with Spring Mills
- **Days:** number of days since the customer was billed
- **Amount:** amount of the bill

Write a program that generates a simple random sample of size 50. How many small companies are in the sample?

Write another program that generates a proportionate stratified random sample of size 50.

#### CONFIDENCE INTERVALS

2. Suppose you are gambling on a roulette wheel. Each time the wheel is spun, the result is one of the outcomes 0, 1, 2, ..., 36. Of these outcomes, 18 are red, 18 are black and one is green. On each spin you bet \$5 that a red outcome will occur and \$1 that the green outcome will occur. If red occurs, you win a net of \$4 (you win \$10 from red and nothing from green). If green occurs, you win a net of \$24 (you win \$30 from green and nothing from red). If black occurs, you lose everything you bet for a loss of \$6.

Write a program that simulates 20 plays from this strategy. Each play should indicate the net amount won or lost. Then, based on these 20 outcomes, find a 95% confidence interval for the total net amount won or lost from 1000 plays of the game. Would you conclude that this strategy is a winning one for you?

3. A market research consultant hired by a leading soft drink company wants to determine the proportion the proportion of consumers who favor its low-calorie brand over the leading low-calorie competitor in a particular geographic region. A random sample of 250 consumers from the market in investigation is provided in **softdrink.xlsx**.

Write a program that finds out a 90% confidence interval for the proportion of all consumers in this market who prefer the company's brand. The file contains the gender and age group for each customer in the sample. Find a separate 90% confidence interval for each gender for the proportion who prefer the company's brand. Then do the same for each age group. Compare the CIs obtained.