Assignment #1, Due 7/5/2024

1. In class we learn there are four types of data. Based on what you can find in National Chengchi University, give your examples of each type of data. Also, show that it is possible to create unreasonable interpretation if falsely categorizing the data type.

2. **Best Paying College Degrees.** Each year America.EDU ranks the best paying college degrees in America. The following data show the median starting salary, the mid-career salary, and the percentage increase from starting salary to mid-career salary for the 20 college degrees with the highest mid-career salary (America.EDU website).

Degree	Starting Salary	Mid-Career Salary	% Increase
Aerospace engineering	59,400	108,000	82
Applied mathematics	56,400	101,000	79
Biomedical engineering	54,800	101,000	84
Chemical engineering	64,800	108,000	67
Civil engineering	53,500	93,400	75
Computer engineering	61,200	87,700	43
Computer science	56,200	97,700	74
Construction management	50,400	87,000	73
Economics	48,800	97,800	100
Electrical engineering	60,800	104,000	71
Finance	47,500	91,500	93
Government	41,500	88,300	113
Information systems	49,300	87,100	77
Management info. systems	50,900	90,300	77
Mathematics	46,400	88,300	90
Nuclear engineering	63,900	104,000	63
Petroleum engineering	93,000	157,000	69
Physics	50,700	99,600	96
Software engineering	56,700	91,300	61
Statistics	50,000	93,400	87

(a) What are the averages of starting and mid-career salary?
(b) Using a class width of 10, construct a histogram for the percentage increase in the starting salary. Comment on the shape of the distribution.
(c) Develop a scatter diagram with the starting salary on the horizontal axis and the percentage increase in the starting salary on the vertical axis. Does there appear to be a relationship between these two variables? Discuss.
(d) What is the sample correlation coefficient between starting salary and mid-career salary? Are there any unusual observations? Discuss.

3. **Multiple-Choice Exam.** Consider a multiple-choice examination with 60 questions. Each question has four possible answers. Assume that a student who has done the homework and attended lectures has probability .7 of answering any question correctly.

   (a) A student must answer 45 or more questions correctly to obtain a grade of A. What percentage of the students who have done their homework and attended lectures will obtain a grade of A on this multiple-choice examination?

   (b) A student who answers 40 to 44 questions correctly will receive a grade of C. What percentage of students who have done their homework and attended lectures will obtain a grade of C on this multiple-choice examination?

   (c) A student must answer 36 or more questions correctly to pass the examination. What percentage of the students who have done their homework and attended lectures will pass the examination?

   (d) Assume that a student has not attended class and has not done the homework for the course. Furthermore, assume that the student will simply guess at the answer to each question. What is the probability that this student will answer 36 or more questions correctly and pass the examination?

4. **Obesity.** Obesity is a risk factor for many health problems such as type 2 diabetes, high blood pressure, joint problems, and gallstones. Using data collected in 2018 through the National Health and Nutrition Examination Survey, the National Institute of Diabetes and Digestive and Kidney Diseases estimates that 37.7% of all adults in the United States have a body mass index (BMI) in excess of 30 and so are categorized as obese. The data in the file Obesity are consistent with these findings.

   (a) Use the Obesity data set to develop a point estimate of the BMI for adults in the United States. Are adults in the United States obese on average?

   (b) What is the sample standard deviation?

   (c) Develop a 95% confidence interval for the BMI of adults in the United States.

5. **U.S. Drivers and Speeding.** ABC News reports that 58% of U.S. drivers admit to speeding. Suppose that a new satellite technology can instantly measure the speed of any vehicle on a U.S. road and determine whether the vehicle is speeding, and

this satellite technology was used to take a random sample of 20,000 vehicles at 6 p.m. EST on a recent Tuesday afternoon.

(a) For this investigation, what is the sampling distribution for sample proportion of vehicles on U.S. roads that speed?

(b) What is the probability that the sample proportion of speeders $\hat{p}$ will be within 1% of the population proportion of speeders?

(c) Suppose the sample proportion of speeders $\hat{p}$ differs from the U.S population proportion of seeders by more than 1%? How would you interpret this result?