DATA ANALYSIS PROJECT 2
A ST 511
FALL 2023

The following is taken from "Case 1 – Medical Malpractice: Descriptive Statistics, Graphics, and Exploratory Data Analysis" by Marlene Smith of the University of Colorado Denver Business School, with slight modifications. (The complete case is provided here: https://www.jmp.com/content/dam/jmp/documents/en/academic/case-study-library/case-study-library-12/business-case-studies/01-medicalmalpractice.pdf.)

Background

According to a recent study published in the *US News and World Report*, the cost of medical malpractice in the United States is $55.6 billion a year, which is 2.4 percent of annual health--care spending. A 2011 study published in the *New England Journal of Medicine* revealed that annually, during the period 1991 to 2005, 7.4 percent of all physicians licensed in the US had a malpractice claim. These staggering numbers not only contribute to the high cost of health care, but the size of successful malpractice claims also contributes to high premiums for medical malpractice insurance. An insurance company wants to develop a better understanding of its claims paid out for medical malpractice lawsuits. Its records show claim payment amounts, as well as information about the presiding physician and the claimant for a number of recently adjudicated or settled lawsuits.

The Data

The data set contains information about the last 118 claim payments made, covering a six-month period. The eight variables in the data set are described below.

Amount	Amount of the claim payment in dollars
Severity	The severity rating of damage to the claimant, from 1 (emotional trauma) to 9 (death)
Age	Age of the claimant in years
Private_Attorney	Whether the claimant was represented by a private attorney (1 = private, 0 = not private)
Marital_Status	Marital status of the claimant (0 = divorced, 1 = single, 2 = married, 3 = widowed, 4 = unknown)
Specialty	Specialty of the physician involved in the lawsuit
Insurance	Type of medical insurance carried by the claimant
Gender	Claimant gender

The Task

Explore the claim payments data and perform analyses to answer questions 1, 2, 3, and 4 below.

1. Describe the distributions of age of the claimant, severity, size of the payment, and whether the claimant was represented by a private attorney.
2. Is there a statistically significant relationship between age of the claimant and size of the payment? If yes, use the slope of the sample regression line to describe the relationship.
3. Is there a statistically significant relationship between severity and size of the payment? If yes, use the slope of the sample regression line to describe the relationship.
4. Does the relationship between severity and size of the payment depend on whether a private attorney was used? (Here, you will want to look at the relationship between severity and size of the payment separately for claimants represented by and not represented by a private attorney.) If yes, how does the relationship differ?

Deliverable

Prepare ***seven slides*** that provide necessary background, describe your statistical methods and results, and discuss the implications of your findings, as follows:

Slide 1: Background
Slide 2: Methods
Slide 3: Question 1 Results
Slide 4: Question 2 Results
Slide 5: Question 3 Results
Slide 6: Question 4 Results
Slide 7: Implications for insurance company

Your audience for this presentation is the insurance company not your statistics professor.

Deliverables are due by 11:59 p.m. on Thursday, December 7, and should be submitted as a single PowerPoint or PDF file.