Group Project

Analysis of Salary Data by Gender

There are many sources of salary data where median salaries for a certain year are published by gender. One source for data on earnings by demographics is the Bureau of Labor Statistics (https://openstax.org/r/bls).

As a group:

- Research salary data by gender for a certain year.
- Compile descriptive statistics where available for statistical measurements including the mean, median, quartiles, and standard deviation by gender.
- Create graphical presentations for this data using histograms, side-by-side box plots, and time series graphs.
- Discuss the following:

a. From your analysis, does there appear to be a wage gap for men's vs. women's salaries?

b. If so, is the wage gap improving over time or worsening over time?

Quantitative Problems

1.

Dataset A	Dat
Number of cars entering a parking garage in a 1-time	Sample of 10 counts: 15, 18, 4, 15, 8, 11, 7, 162

a. For Dataset A, calculate the mean and a 10% trimmed mean (round answers to 1 decimal place). Use technology as appropriate.

b. Is there any benefit in using the trimmed mean versus the mean for this dataset?

2.

Dataset B Description	Data
Weights of packages on a delivery truck (weights in pounds)	Sample of 12 packages: 3.2, 9.1, 4.3, 1.9, 2.6, 4.7, 0.7, 1.3, 3.9, 6.4, 5.0, 25.4

a. For Dataset B, calculate the mean and median (round answers to 1 decimal place). Use technology as appropriate.

b. Is there any preference in using the mean versus the median for this dataset?

c. If so, would the mean or the median be preferred as a measure of central tendency for this dataset?

3. Dataset C Description Data

Amount of time for a pain reliever to provide relief from migraine headache (time in minutes)	Sample of 10 patients: 12.1, 13.8, 9.4, 15.9, 11.5, 14.6, 18.1, 12.7, 11.0, 14.2

a. For Dataset C, calculate the range, standard deviation, and variance (round answers to 1 decimal place). Use technology as appropriate and include appropriate units as part of your answers.

b. What does the standard deviation of this dataset communicate regarding the amount of time to provide relief from a migraine headache?

c. Would a medical researcher prefer to use the standard deviation or the variation for this dataset, and why?

4. Dataset D Description Data

Credit card balances for consumers (in dollars)	Sample of 9 consumers: $878.32, $671.54, $990.35, $801.85, $4303.76, $1285.03, $289.44, $78.61, $952.83

a. For Dataset D, calculate the quartiles of the dataset.

b. For this same dataset, calculate the interquartile range (IQR).

c. Determine if there are any potential outliers in this dataset.

5. Dataset E Description Data

Salaries for employees at a technology company (in dollars)	Sample of 10 employees: $72,500; $84,300; $58,900; $55,100; $82,000; $94,600; $82,200; $104,600; $63,800; $77,500

For Dataset E, determine the percentile for the employee with a salary of $82,000. Round your answer to

the nearest whole number percentile.

6. At a website development company, the average number of sick days taken by employees per year is 9 days with a standard deviation of 2.3 days.

a. A certain employee has taken 2 sick days for the year. Calculate the z-score for this employee (round your answer to 2 decimal places).

b. What does the z-score communicate about this employee's number of sick days as compared to other employees in the company?

c. A different employee has taken 17 sick days for the year. Calculate the z-score for this employee (round your answer to 2 decimal places).

d. What does the z-score communicate about this employee's number of sick days as compared to other employees in the company?

7. Two cards are selected at random from a standard 52-card deck. Find the probability that the second card is an Ace, given that the first card selected was an Ace (assume the first card is not replaced before selecting the second card). Round your answer to 4 decimal places.

8. From past data, it's known that the probability that an individual owns a dog is 0.42. Three random people are asked if they own a dog. (Round answers to 3 decimal places). a. Find the probability that all three people own a dog.

b. Find the probability that none of the three people own a dog.

c. Find the probability that at least one of the people owns a dog.

9. In a group of 100 people, 35 people wear glasses, 40 people wear contacts, 12 wear both glasses and contacts, and 25 wear neither glasses nor contacts.

a. Are the events "person wears glasses" and "person wears contacts" mutually exclusive?

b. Calculate the probability that a randomly chosen person wears glasses or contacts. Round your answer to 3 decimal places.

10. A medical researcher collects statistical data as follows:

- 23% of senior citizens (65 or older) get the flu each year.
- 32% of people under 65 get the flu each year.
- In the population, there are 15% senior citizens.

a. Are the events "being a senior" and "getting the flu" dependent or independent?

b. Find the probability that a person selected at random is a senior and will get the flu.

c. Find the probability that a person selected at random is under 65 years old and will get the flu.

11. A certain type of COVID-related virus affects 5% of the population. A screening test is available to detect the virus. Assume the probability of a true positive from the screening test is 85%, meaning that the probability that a person has a positive test result, given that they actually have the virus, is 0.85. Also assume the probability of a false positive from the screening test is 12%, which indicates the probability that a person has a positive test result, given that they do not have the virus, is 0.12. Calculate the probability that a person actually has the virus given that the screening test shows a positive result. Round your answer to 3 decimal places.

12. You are taking a multiple-choice test with 10 questions where each question has 5 possible answers: (a), (b), (c), (d), and (e). You did not study, so you will guess at each question. What is the probability of getting exactly 5 questions correct?

13. The average arrival rate of trucks at a truck loading station is 15 per day. Calculate the probability that 17 trucks arrive at the loading station on a particular day.

14. In a certain city, cell phone bills are known to follow a normal distribution with a mean of $92 and standard deviation of $14.

3 • Quantitative Problems

a. Find the probability that a randomly selected cell phone customer has a monthly bill less than $75.

Find the probability that a randomly selected cell phone customer has a monthly bill more than $85.