Assignment - 10

Pradyoth Singenahalli Prabhu - 02071847

Problem-1

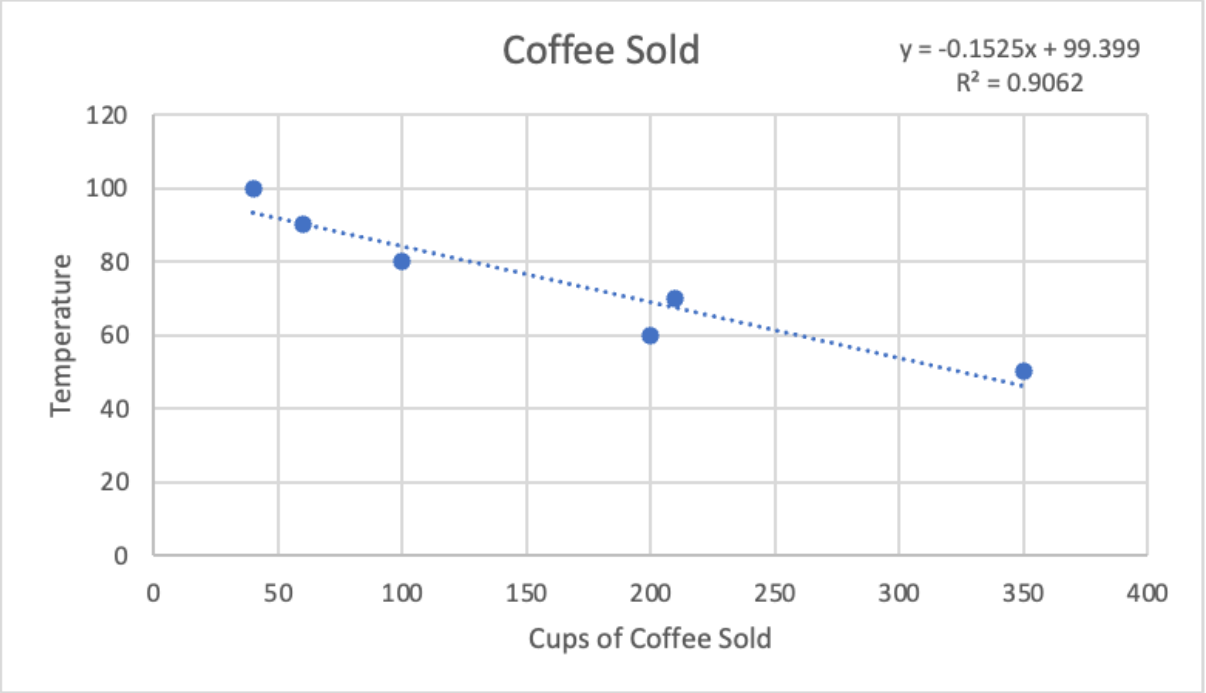
Max believes that the sales of coffee at his coffee shop depend upon the weather. He has taken a sample for 5 days. Below you are given the results of the sample.

Cups of Coffee Sold	Temperature
350	50
200	60
210	70
100	80
60	90
40	100

a. Which variable is the dependent variable?

In the given scenario, the dependent variable is the **"Cups of Coffee Sold”**.

b. Compute the least squares estimated line.



y = -5.9429x + 605.71

c. Compute the correlation coefficient between temperature and the sales of coffee.

R^2 = **0.9062**

r_{xy} = **0.951945377**

d. Predict sales of a 90-degree day.

$\hat{y}$ = (-5.9429 * 90) + 605.71

$\hat{y}$ = **70.849**

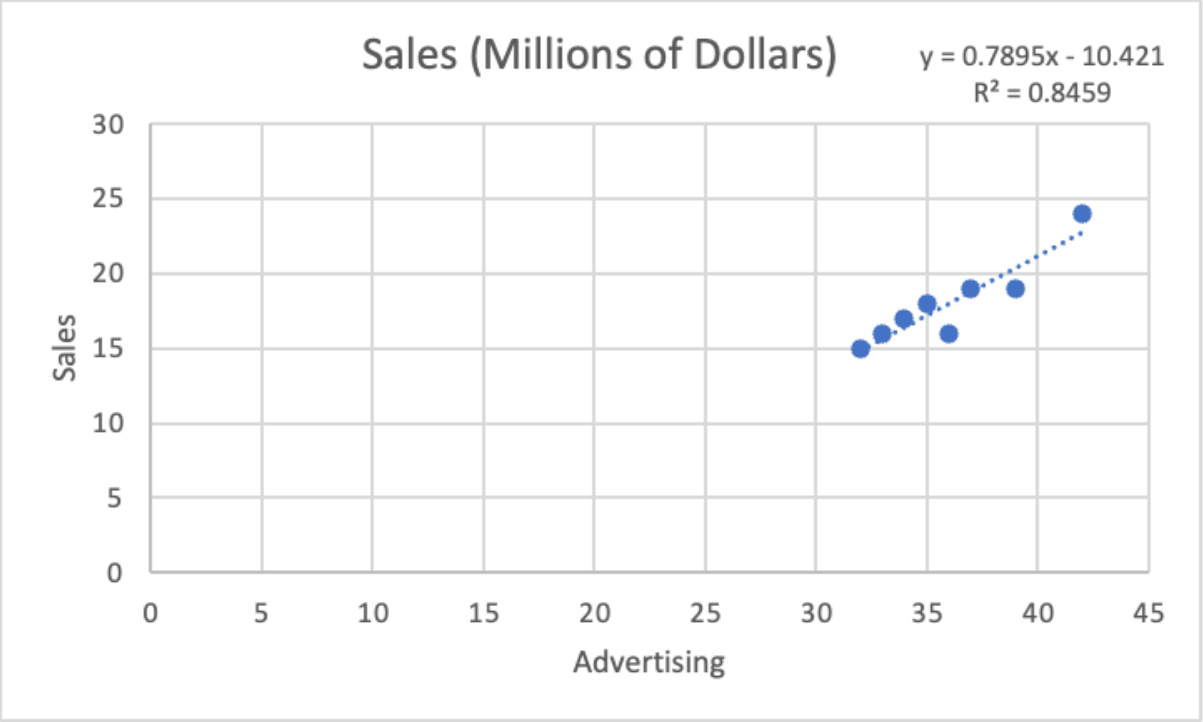
Problem - 2

The following data represent a company's yearly sales volume and its advertising expenditure over a period of 8 years.

Year	Sales (Millions of Dollars)	Advertising ($10,000)
1994	15	32
1995	16	33
1996	18	35

Year	Sales (Millions of Dollars)	Advertising ($10,000)
1997	17	34
1998	16	36
1999	19	37
2000	19	39
2001	24	42

a. Develop a scatter diagram of sales versus advertising.



b. Use the least-squares method to compute an estimated regression line between sales and advertising.

b1	0.789473684
b0	-10.42105263

y = 0.789473684*x - 10.42105263

c. If the company's advertising expenditure is $400,000, what are the predicted sales?

$\hat{y}$ `=B14*40 + B15` = **21.15789474**

d. What does the slope of the estimated regression line indicate?

The slope of the estimated regression line signifies the rate of change in sales for every $10,000 increment in advertising expenditure. In this context, a slope of 0.789 suggests that for every additional $10,000 spent on advertising, the model predicts an increase in sales by $789,000. This positive slope indicates a positive correlation, implying that higher advertising spending is associated with higher sales.

In practical terms, if the advertising budget increases by $10,000, the model forecasts a growth in sales by $789,000. The numeric value of the slope, 0.789, quantifies the change in sales in dollars for each ten-thousand-dollar increase in advertising expenditure.

The positive gradient of the slope aligns with the understanding that an uptick in advertising investment is anticipated to result in a corresponding growth in sales. This insight is derived from the model's analysis of the data, suggesting a positive relationship between advertising spending and sales outcomes.

e. Compute the coefficient of determination and fully interpret its meaning.

$R^2 = 0.8459$

The coefficient of determination (R^2) is calculated to be 0.8459. This value indicates the proportion of the variance in sales that can be explained by the linear regression model based on advertising expenditure. As R^2 approaches 1, it signifies that a larger proportion of the variability in sales can be accounted for by the model.

In this context, an R^2 of 0.8459 suggests a strong connection between advertising expenditure and the volume of goods sold by a company. The higher the R^2 value, the more accurate the linear regression model is in predicting sales outcomes using the provided data.

Therefore, with an R^2 value of 0.8459, it is reasonable to conclude that there is a robust and reliable linear relationship between advertising expenditure and sales,

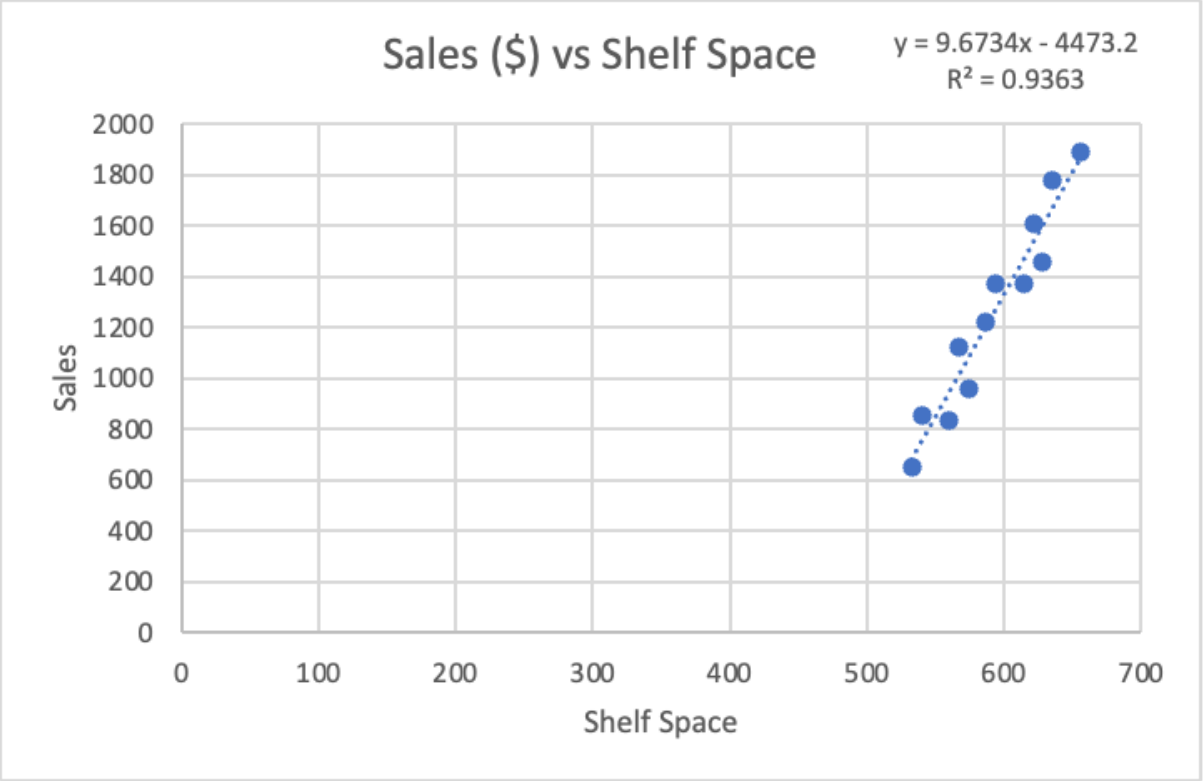
indicating that approximately 84.59% of the variability in sales can be explained by changes in advertising spending.

Problem - 3

A market research analyst for a brand of cereal is interested in finding out if there is a relationship between the sales generated and the shelf space used to display the cereal. She conducted a study and collected data from 12 different stores selling this brand of cereal.

Shelf Space (Sq in)	Sales ($)
574	960
635	1779
533	651
560	831
628	1460
615	1370
540	851
587	1220
656	1889
594	1370
622	1609
567	1120

The data contains sales $ generated for a certain month and the shelf space dedicated to the product. Analyze this data using the appropriate method (Compute regression equation and the coefficient of determination).



y = 9.6734x - 4473.2

R^2 = **0.9363**

coefficient of determination = **93.63%**

Problem - 4

The following data show the brand, price ($), and overall score for stereo headphones that were tested by Consumer Reports. The overall score is based on sound quality and the effectiveness of ambient noise reduction. Scores range from (lowest) to (highest).

Brand	Price	Score
Bose	180	76
Skullcandy	150	71
Koss	95	61
Phillips/O'Neill	70	56
Denon	70	40
JVC	35	26

(a) Compute the estimated regression equation.

y = 0.3181x + 23.194

(b) Compute SST, SSR, and SSE. (Three decimal places).

SST	1800
SSR	1512.755
SSE	287.245

(c) Compute the coefficient of determination. (Three decimal places).

$R^2 = \frac{SSR}{SST}$

R^2 = 0.840

(d) What is the value of the sample correlation coefficient? (Three decimal places).

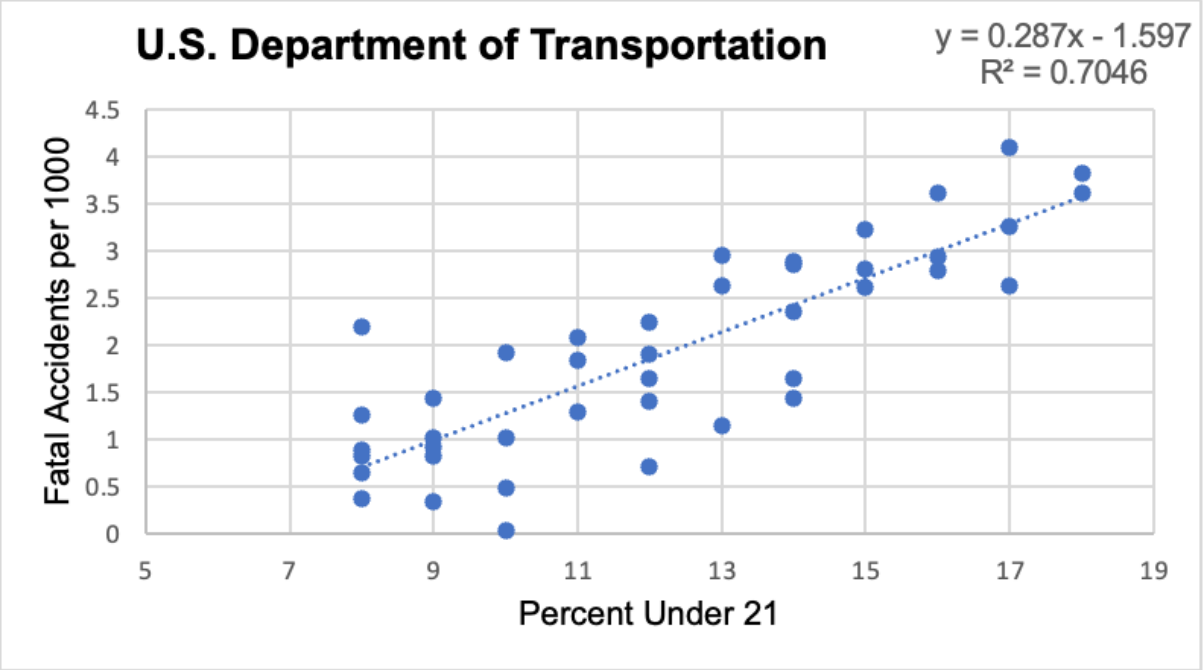
Correlation co-efficient	0.9167439

Case-Study

As part of a study on transportation safety, the U.S. Department of Transportation collected data on the number of fatal accidents per 1000 licenses and the percentage of licensed3 drivers under the age of 21 in a sample of 42 cities. Data collected over one year follow. These data are contained in the file ***Safety***.

a) Develop numerical and graphical summaries of the data.

Percent Under 21		Fatal Accidents per 1000	
Mean	12.262	Mean	1.922
Standard Error	0.483	Standard Error	0.165
Median	12	Median	1.881
Mode	8	Mode	#N/A
Standard Deviation	3.132	Standard Deviation	1.071
Sample Variance	9.808	Sample Variance	1.147
Kurtosis	-1.137	Kurtosis	-0.975
Skewness	0.210	Skewness	0.193
Range	10	Range	4.061
Minimum	8	Minimum	0.039
Maximum	18	Maximum	4.1
Sum	515	Sum	80.741
Count	42	Count	42



The numerical and graphical summaries underscore the need to understand the variations in both the demographic composition of drivers and the associated accident rates among different cities. The scatter plot reinforces the findings of a **positive linear relationship** between the percentage of drivers under 21 and fatal accidents, supporting the conclusions drawn from the linear regression analysis.

b) Use regression analysis to investigate the relationship between the number of fatal accidents and the percentage of drivers under the age of 21. Discuss your findings.

This summary explores the potential relationship between the percentage of drivers under 21 and the number of fatal road accidents per 1000 driving licenses, based on a regression analysis without assuming any direction or strength of correlation.

Findings:

- **Multiple R:** The value of 0.8394 indicates a statistically significant correlation between the two variables. However, it only tells us that there is a **linear relationship**, not necessarily that one variable causes the other to change.
- **R^2:** The value of 0.7046 signifies that 70.46% of the variation in fatal accidents can be explained by changes in the percentage of drivers under 21. This highlights the potential importance of this variable in predicting crash rates, but does not confirm it as the sole contributor.

- **Regression Coefficients:**
  - **Intercept:** The intercept of -1.597 represents the predicted average number of fatal accidents per 1000 licenses when there are **no drivers under 21**. Its meaning in this context is unclear and should be interpreted with caution.
  - **"Percent under twenty-one" coefficient:** This value of 0.287 indicates that for every **one unit increase** in the percentage of young drivers, there is an **estimated increase of 0.287** fatal accidents per 1000 licenses. This quantifies the association between the variables, but only in terms of **linear trends**, not necessarily causation.

Conclusion:

This analysis establishes a statistically significant, **positive linear relationship** between the percentage of drivers under 21 and the number of fatal road accidents. However, it is important to remember that:

- This relationship does not imply **causation**. Other factors could also be influencing both variables.
- The direction of the relationship is determined by the **sign** of the coefficient, not by the order of variables in the regression.
- The R^2 value tells us how much **variance** is explained by the model, but not how accurately it can **predict** individual cases.

c) What conclusion and recommendations can you derive from your analysis?

Based on the regression analysis, the following considerations and potential actions could be explored:

1. Relationship between Young Drivers and Fatal Accidents:

- The analysis suggests a statistically significant linear association between the percentage of drivers under 21 and the number of fatal road accidents. It explains approximately 70.5% of the variance in fatal crashes, highlighting the potential influence of this variable.
- However, it is important to remember that correlation does not necessarily imply causation. Other factors could also be influencing both variables, and the direction of the relationship cannot be determined from this analysis alone.

2. Potential Actions:

- **Education and Awareness:** Implementing targeted educational programs and awareness campaigns for young drivers could lead to improved knowledge about safe driving practices, road hazards, and the consequences of risky behavior. This may contribute to better decision-making on the road.
- **Community Engagement:** Engaging in community-wide initiatives to raise awareness about road safety could involve educational campaigns, workshops, and collaborations with local organizations. This can foster a culture of responsible driving and community support for safety measures.
- **Policy Considerations:** Exploring the potential effectiveness of special rules or guidelines for young drivers, such as graduated licensing programs, requires further investigation. These programs could gradually introduce driving privileges as individuals gain experience and demonstrate responsible behavior.
- **Stakeholder Collaboration:** Collaborating with various stakeholders, including government agencies, law enforcement, educational institutions, and community organizations, can enhance the effectiveness of road safety interventions. Partnerships can facilitate knowledge sharing, resource allocation, and implementation of coordinated strategies.
- **Evaluation and Monitoring:** Conduct regular assessments of implemented road safety measures to evaluate their impact. Measuring effectiveness and identifying areas for improvement through data analysis and stakeholder feedback is crucial for a dynamic approach to road safety.
- **Continuous Improvement:** Establishing a framework for continuous improvement in road safety measures involves ongoing monitoring, feedback mechanisms, and adjustments to strategies based on evolving data and circumstances. This ensures adaptation to changing trends and optimizes the effectiveness of interventions.
- **Age Considerations:** Acknowledging the importance of considering age demographics in road safety policies and interventions is crucial. Understanding the specific risks and challenges associated with different age groups allows for targeted and effective measures.
- **Further Research:** While the analysis provides valuable insights, continued exploration and research into the factors influencing road safety is needed. Ongoing studies can contribute to a deeper understanding of the dynamics involved and inform evidence-based policymaking.

- **Data Considerations:** Any policy changes based on demographic considerations should be informed by new and relevant data. Without updated data, such efforts may be less impactful or even ineffective.

In summary, the analysis encourages exploration of various initiatives and interventions to address the potential relationship between young drivers and fatal accidents. Recognizing the need for further research, collaboration, and ongoing evaluation can contribute to a comprehensive and adaptable road safety strategy.