STATISTICS FOR COMPUTING

Higher National Diploma in Information Systems for Management

24.1F

KAHNDISM24.1F-010

KAHNDISM24.1F-008

KAHNDISM24.1F-O11

School of Computing and Engineering

National Institute of Business Management

Table of Content

1.Introduction

We will be discussing an important public health topic—the percentage of people who smoke within our city. By analyzing the smoking trends in our local population, we can gain insights into the impact of smoking on public health, explore factors contributing to these habits, and consider potential measures for reducing smoking rates. Let's dive into the data and see what the numbers tell us.

2.Data Exploration and Cleaning

- **Variables Description;**

    The dataset contains 200 entries. The columns represent age,

    gender, income level, smoking consumption frequency.

- **Key Variables;**

    Independant Variables-Age, Gender, Income Level, Types of smoking

    Dependent Variables-Smoking consumption frequency

- **Data Cleaning;**

    There are no missing values in the dataset

3.Descriptive Statistics

Descriptive statistics for relevant variables

For this section, we calculated key descriptive statistics for smoking usage of kandy city. These statistics include the mean, median, standard deviation and more.

Statistics

		age	gender	income	types of smoking
N	Valid	200	200	200	200
	Missing	0	0	0	0
Mean		29.93	1.11	3.43	2.96
Median		29.00	1.00	3.00	3.00
Mode		33	1	3	1
Std. Deviation		7.061	.314	1.321	1.716
Variance		49.859	.098	1.744	2.943
Range		27	1	5	6
Minimum		18	1	1	1
Maximum		45	2	6	7

age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18	5	2.5	2.5	2.5
	19	8	4.0	4.0	6.5
	20	7	3.5	3.5	10.0
	21	5	2.5	2.5	12.5
	22	12	6.0	6.0	18.5
	23	6	3.0	3.0	21.5
	24	6	3.0	3.0	24.5
	25	8	4.0	4.0	28.5
	26	14	7.0	7.0	35.5
	27	10	5.0	5.0	40.5
	28	13	6.5	6.5	47.0
	29	12	6.0	6.0	53.0
	30	8	4.0	4.0	57.0
	31	2	1.0	1.0	58.0
	32	6	3.0	3.0	61.0
	33	15	7.5	7.5	68.5
	34	10	5.0	5.0	73.5
	35	4	2.0	2.0	75.5
	36	8	4.0	4.0	79.5
	37	5	2.5	2.5	82.0
	38	10	5.0	5.0	87.0
	39	7	3.5	3.5	90.5
	40	3	1.5	1.5	92.0
	41	3	1.5	1.5	93.5
	42	1	.5	.5	94.0
	43	6	3.0	3.0	97.0
	44	2	1.0	1.0	98.0
	45	4	2.0	2.0	100.0
	Total	200	100.0	100.0	

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	178	89.0	89.0	89.0
	female	22	11.0	11.0	100.0
	Total	200	100.0	100.0	

income

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Less than Rs.1000	16	8.0	8.0	8.0
	Rs.1000-2500	32	16.0	16.0	24.0
	Rs. 2500-5000	58	29.0	29.0	53.0
	Rs. 5000-10000	51	25.5	25.5	78.5
	More than Rs.10000	30	15.0	15.0	93.5
	Not applicable	13	6.5	6.5	100.0
	Total	200	100.0	100.0	

types of smoking

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cigarettes	56	28.0	28.0	28.0
	Beedis	32	16.0	16.0	44.0
	Cigars	41	20.5	20.5	64.5
	Chewing tobacco	31	15.5	15.5	80.0
	E-cigarettes/Vapes	22	11.0	11.0	91.0
	Other	11	5.5	5.5	96.5
	I don't use tobacco products	7	3.5	3.5	100.0
	Total	200	100.0	100.0	

Case Processing Summary

		Cases							
		Valid		Missing		Total			
	gender	N	Percent	N	Percent	N	Percent		
age	male	178	100.0%	0	0.0%	178	100.0%		
	female	22	100.0%	0	0.0%	22	100.0%		

Histogram

8
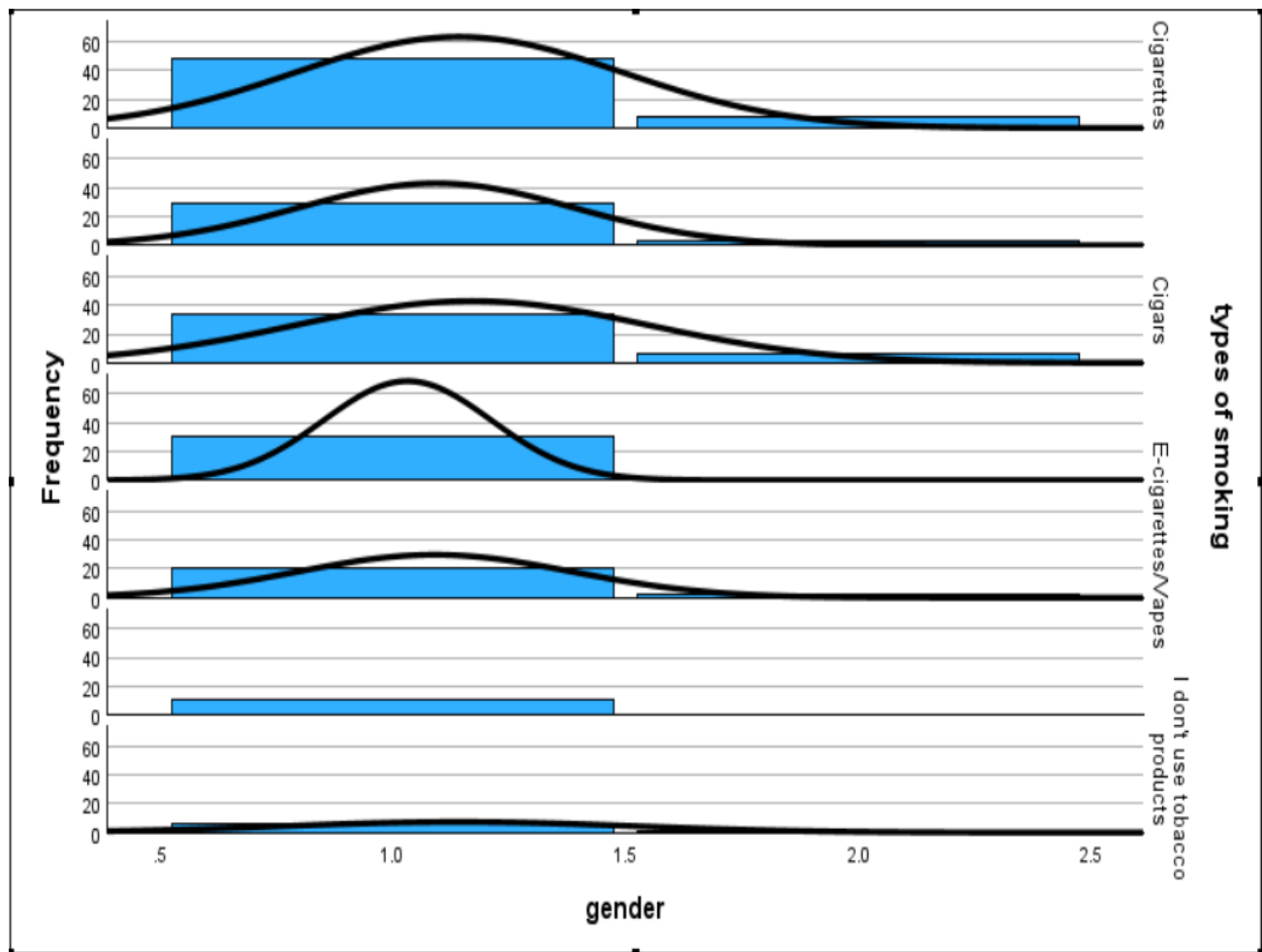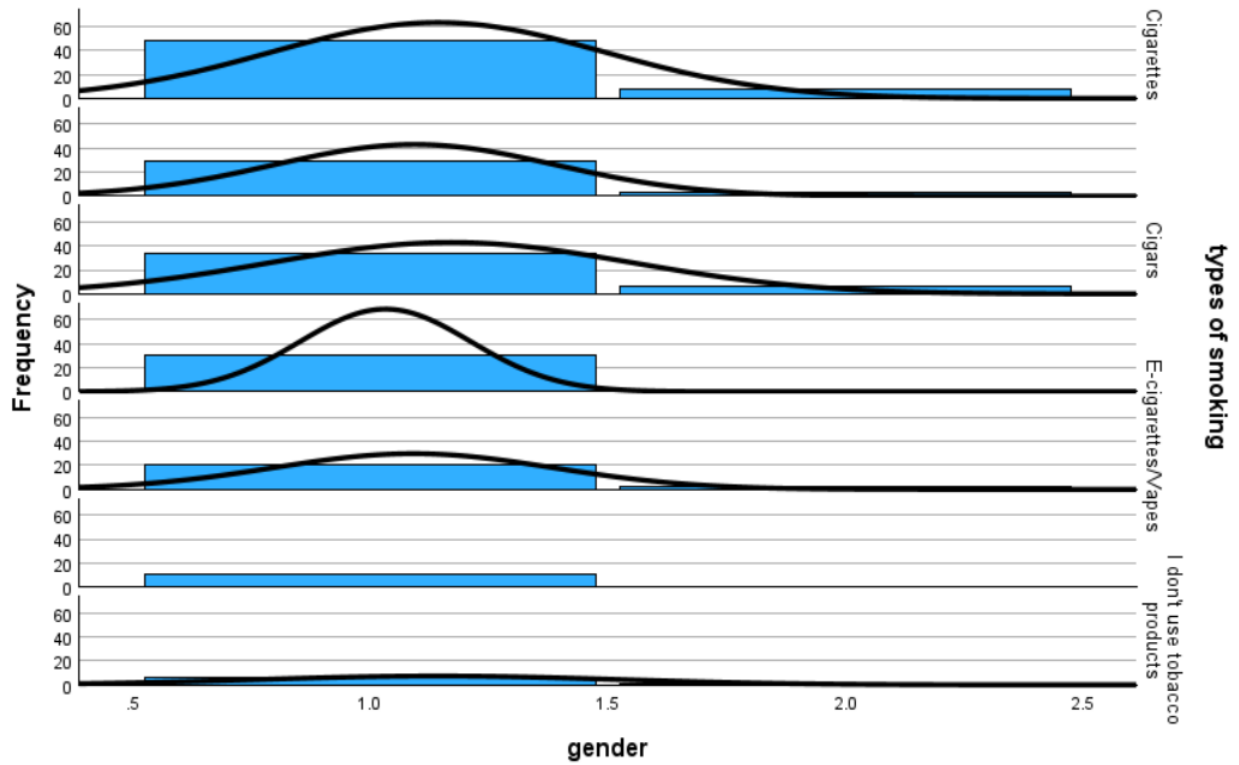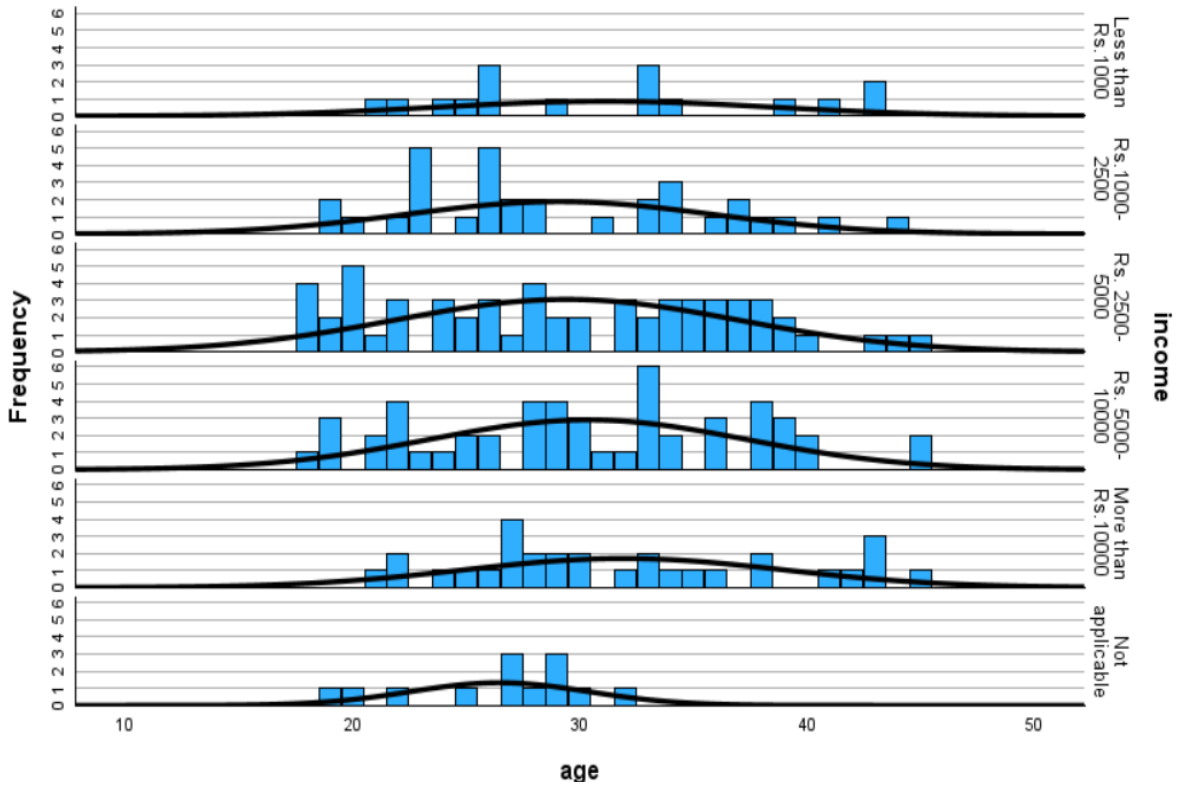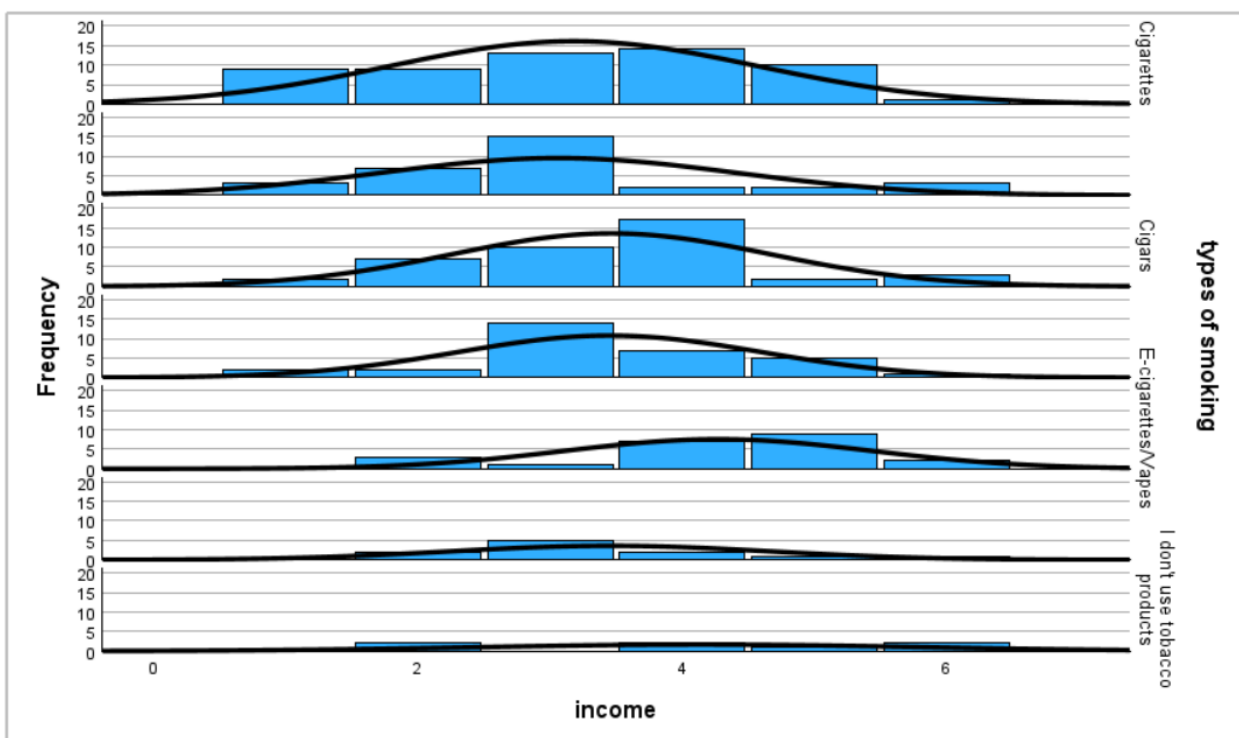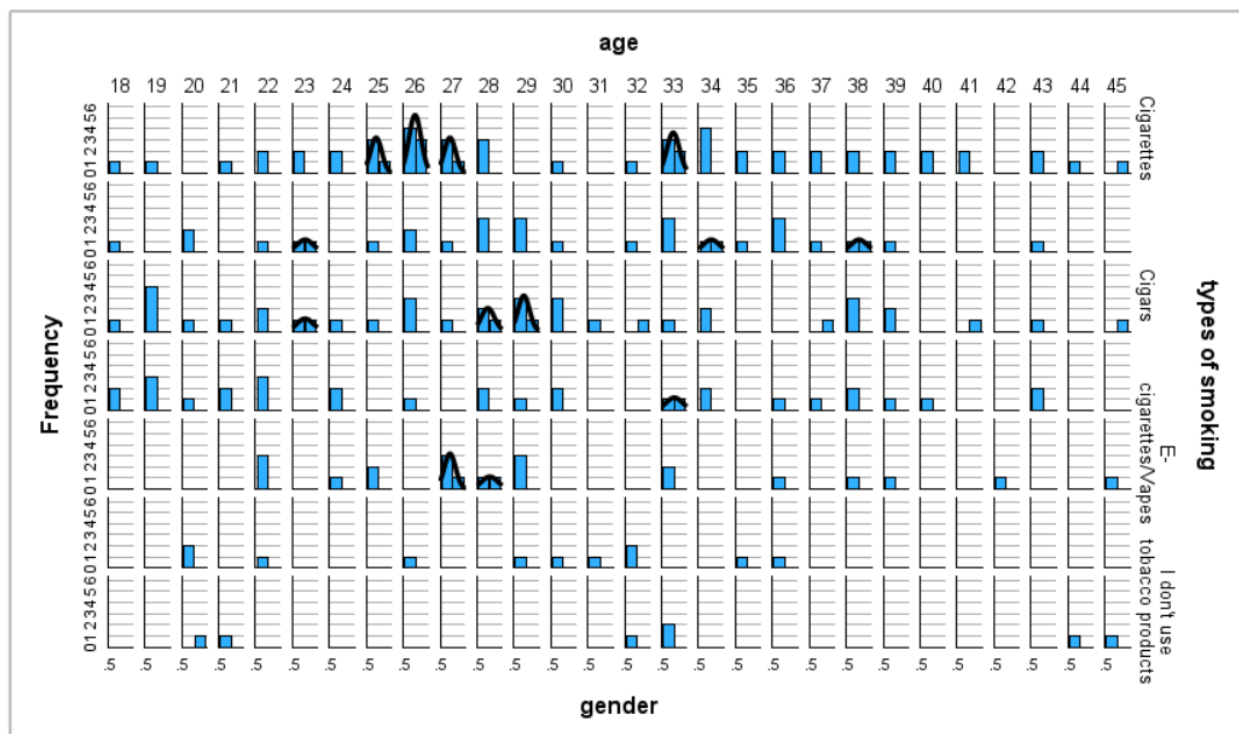
4. Relationship Between Variables

Independent Variables (Factors that might influence smoking consumption frequency):

1. Age: Different age groups might have varying patterns of smoking consumption.

2. Gender: Gender differences may impact smoking habits, frequency, or social acceptance of smoking use.

3. Income Level: Economic status might influence the ability to purchase smoking and consumption habits.

Dependent Variable (The outcome being measured):

• Smoking Consumption Frequency

Relationship between Variables:

• Age and Smoking Consumption Frequency: You would analyze if certain age groups (e.g., younger vs. older adults) tend to consume smoking more or less frequently.

• Gender and Smoking Consumption Frequency: This investigates whether men or women consume smoking more frequently.

• Income Level and smoking Consumption Frequency: Examining if individuals with higher or lower income levels tend to consume smoking more often.
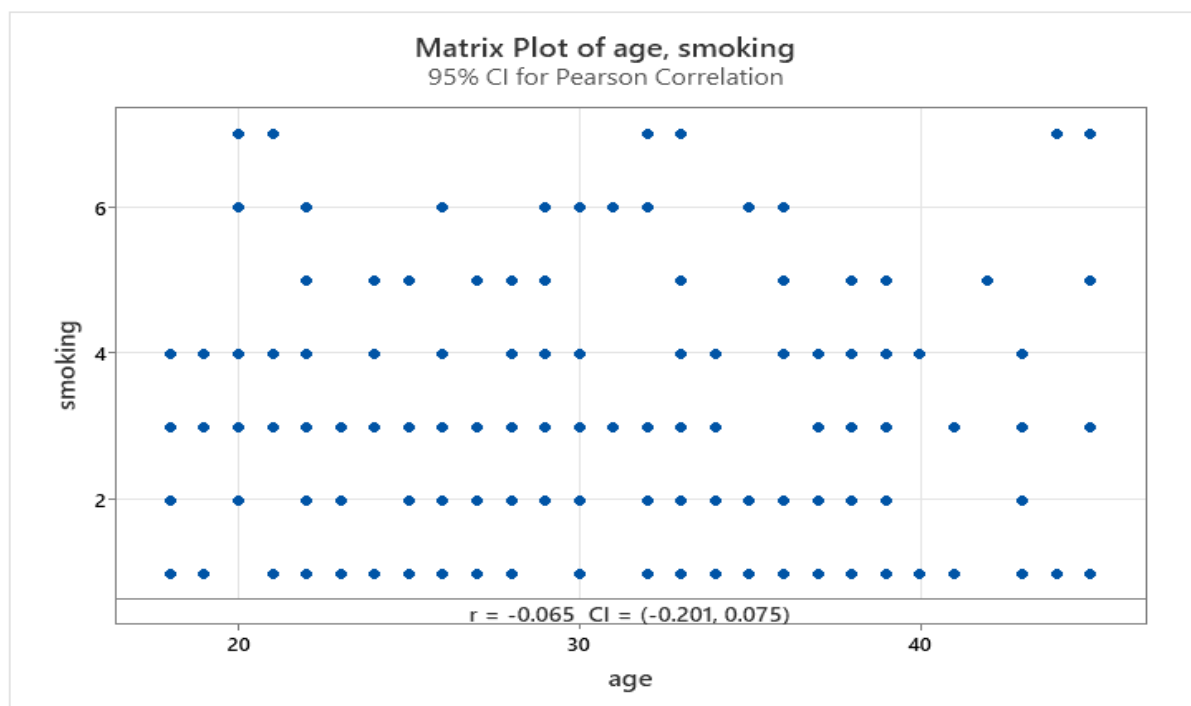
5.Methodology

Correlation Analysis

- Age vs Smoking

Correlations

	age
smoking	-0.065

The graph shows a matrix plot of age and smoking. The x-axis represents age, and the y-axis represents smoking. Each dot in the plot represents a data point, with the x-coordinate corresponding to the age of the individual and the y-coordinate corresponding to their smoking habit.

 The plot also includes a Pearson correlation coefficient (r) of -0.065, with a 95% confidence interval of (-0.201, 0.075). This indicates that there is a weak negative correlation between age and smoking. In other words, as age increases, there is a slight tendency for smoking to decrease. However, this correlation is not statistically significant, as the confidence interval includes zero.
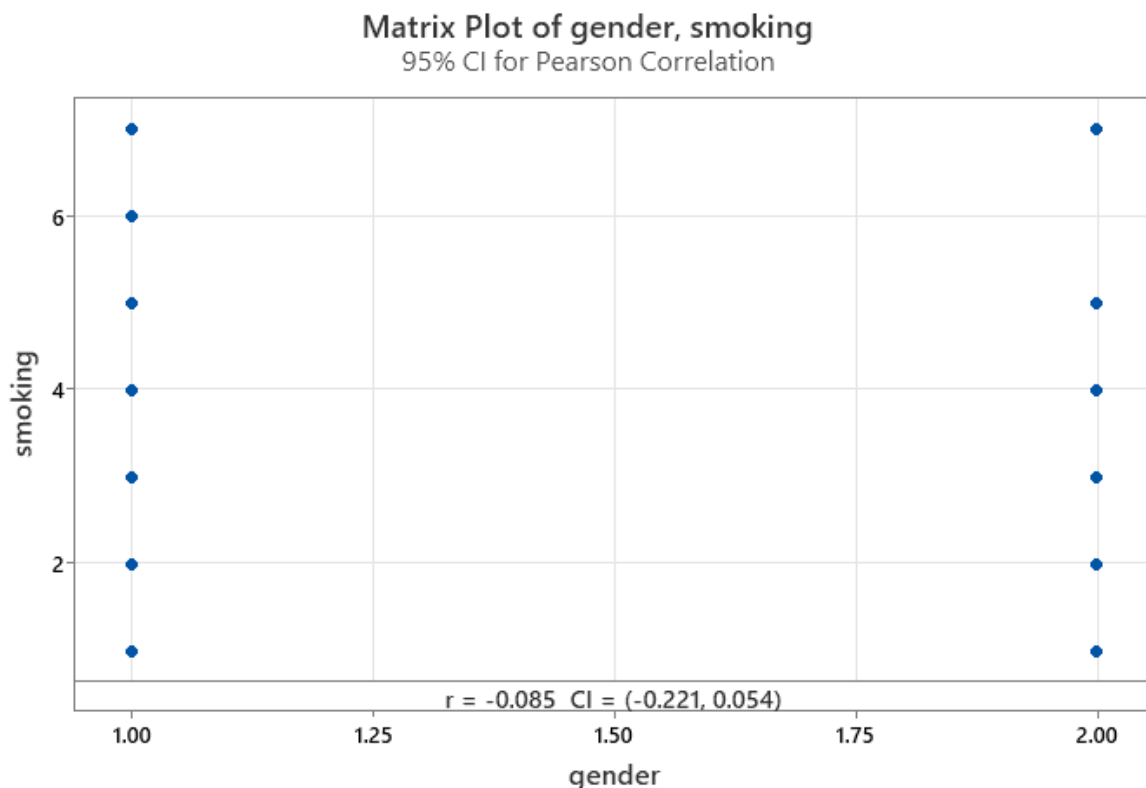

Matrix Plot of age, smoking
95% CI for Pearson Correlation
r = -0.065 CI = (-0.201, 0.075)

- Gender vs Smoking

Correlations

	gender
smoking	-0.085

The graph shows a matrix plot of gender and smoking. The x-axis represents gender, and the y-axis represents smoking. Each dot in the plot represents a data point, with the x-coordinate corresponding to the gender of the individual and the y-coordinate corresponding to their smoking habit.

The plot also includes a Pearson correlation coefficient (r) of -0.085, with a 95% confidence interval of (-0.221, 0.054). This indicates that there is a weak negative correlation between gender and smoking. In other words, males tend to smoke slightly less than females, but this correlation is not statistically significant, as the confidence interval includes zero.


Matrix Plot of gender, smoking
95% CI for Pearson Correlation

r = -0.085 CI = (-0.221, 0.054)

- Age vs Income

Correlations

	age
income	-0.002

The graph shows a matrix plot of age and income. The x-axis represents age, and the y-axis represents income. Each dot in the plot represents a data point, with the x-coordinate corresponding to the age of the individual and the y-coordinate corresponding to their income.

The plot also includes a Pearson correlation coefficient (r) of -0.002, with a 95% confidence interval of (-0.141, 0.136). This indicates that there is a negligible correlation between age and income. In other words, there is no significant relationship between age and income in the data.
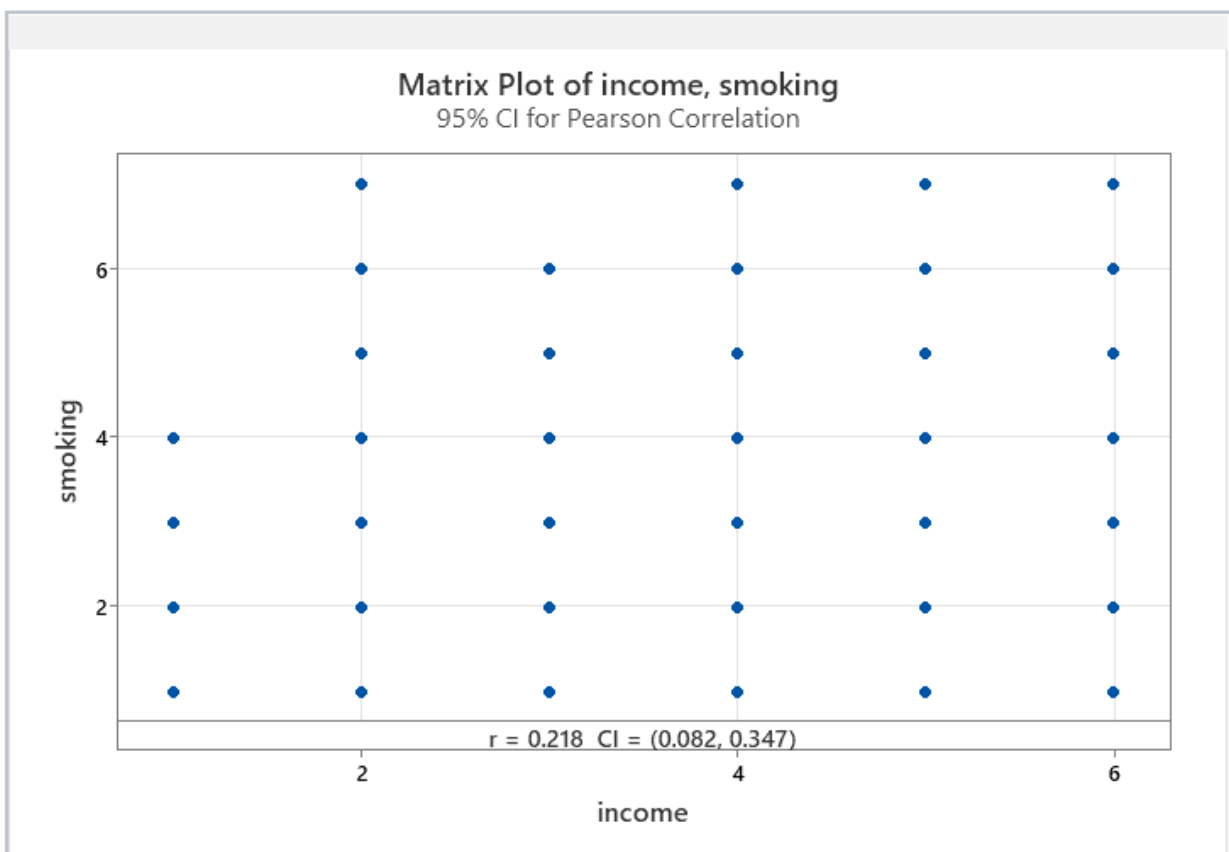

Matrix Plot of age, income
95% CI for Pearson Correlation
r = -0.002 CI = (-0.141, 0.136)
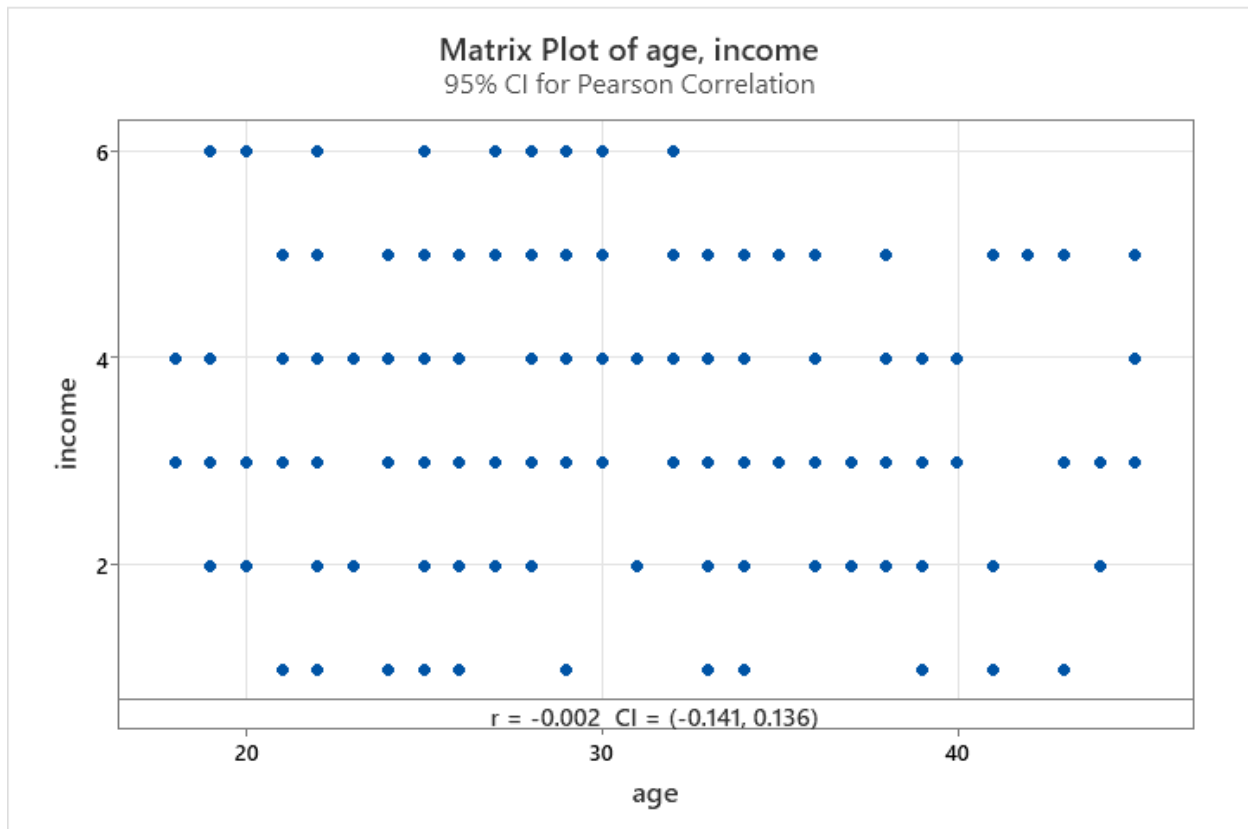
- Income vs Smoking

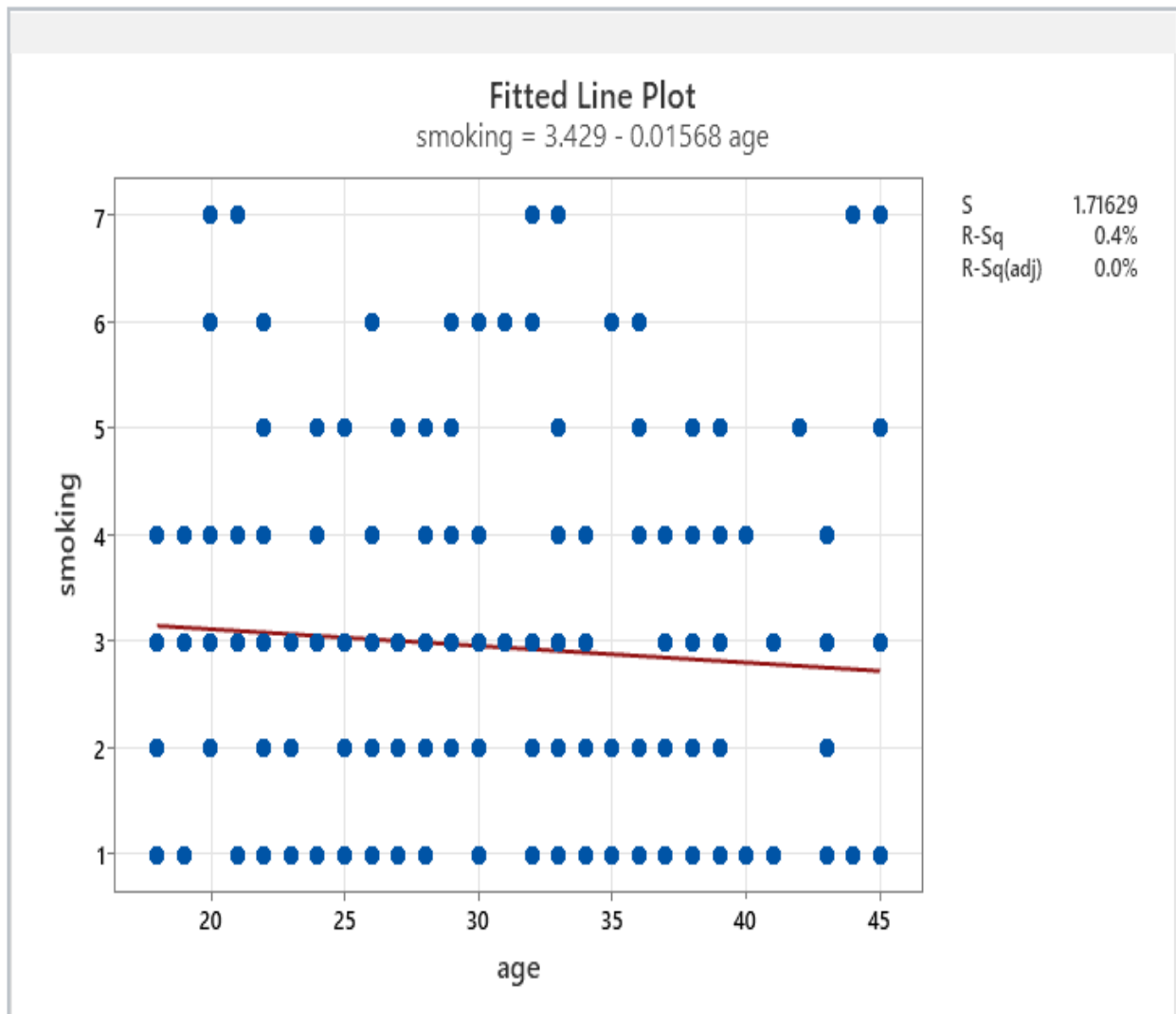## Correlations

	income
smoking	0.218

The graph shows a matrix plot of income and smoking. The x-axis represents income, and the y-axis represents smoking. Each dot in the plot represents a data point, with the x-coordinate corresponding to the income of the individual and the y-coordinate corresponding to their smoking habit.

The plot also includes a Pearson correlation coefficient (r) of 0.218, with a 95% confidence interval of (0.082, 0.347). This indicates that there is a weak positive correlation between income and smoking. In other words, as income increases, there is a slight tendency for smoking to increase as well. However, this correlation is not very strong, and the confidence interval includes zero, suggesting that the relationship is not statistically significant.
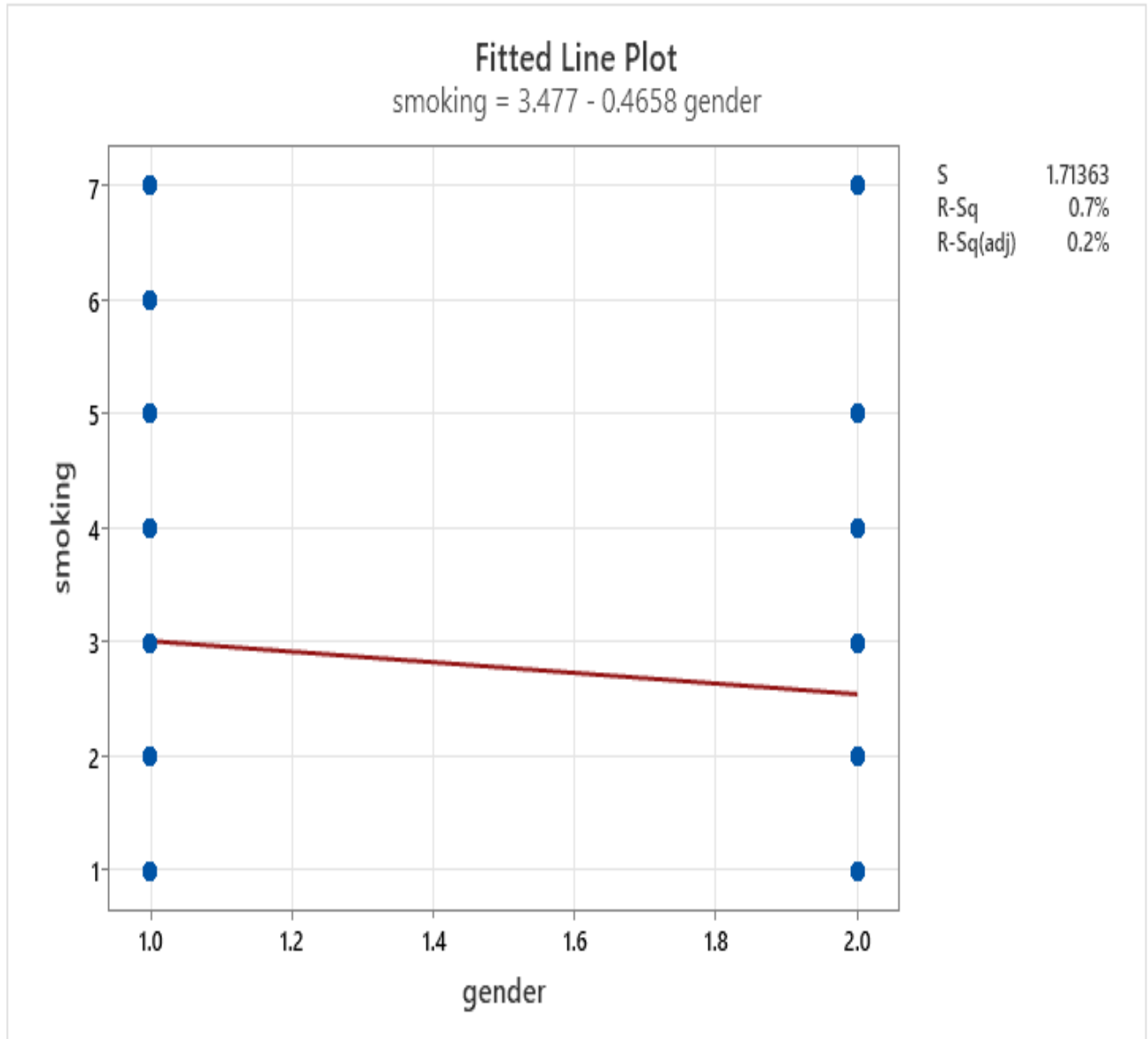

Matrix Plot of income, smoking
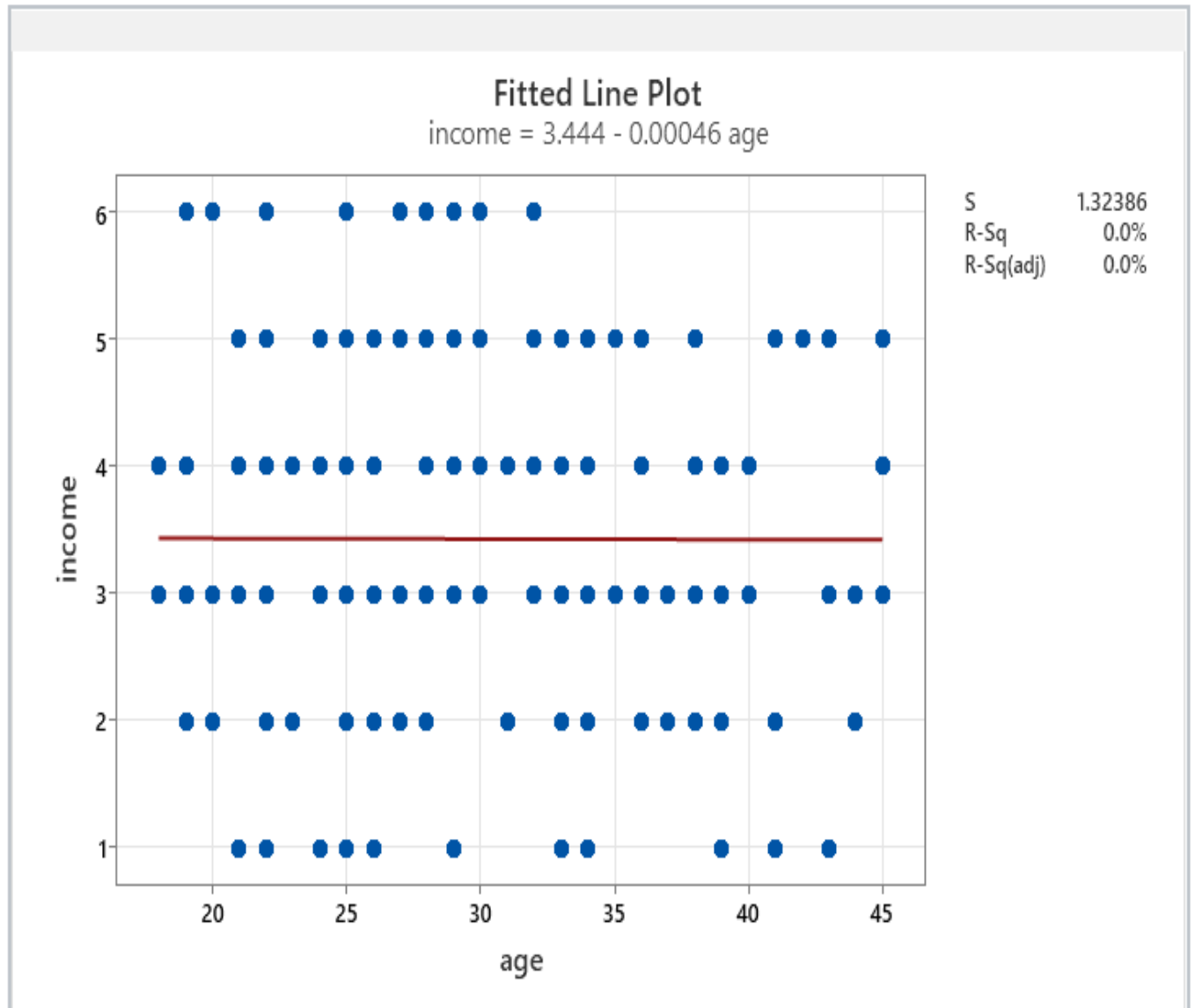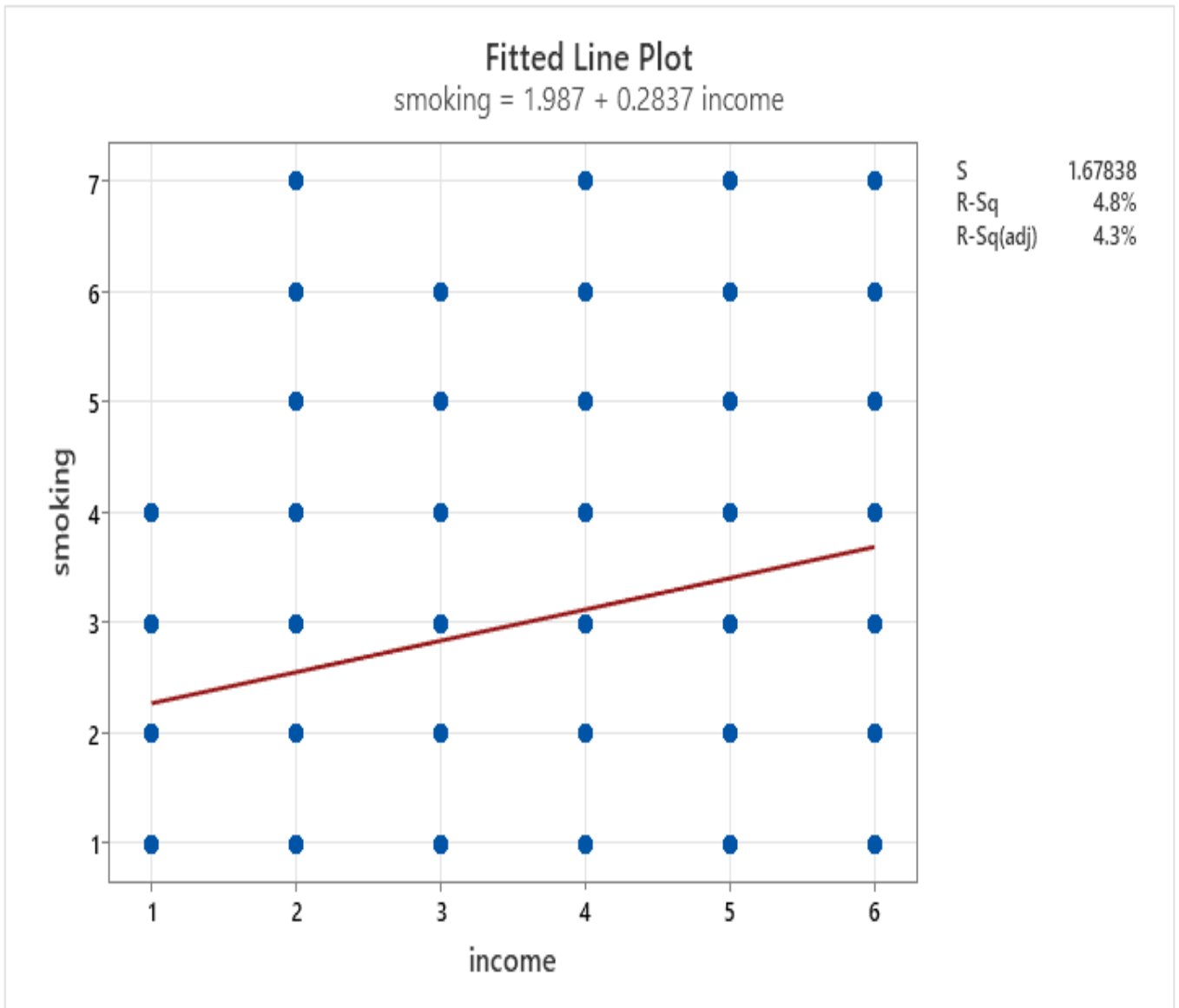95% CI for Pearson Correlation
r = 0.218 CI = (0.082, 0.347)

Regression Analysis

- Age vs Smoking

Fitted Line Plot
smoking = 3.429 - 0.01568 age

S	1.71629
R-Sq	0.4%
R-Sq(adj)	0.0%

- Gender vs Smoking

Fitted Line Plot
smoking = 3.477 - 0.4658 gender

S	1.71363
R-Sq	0.7%
R-Sq(adj)	0.2%

- Age vs Income

- Income vs Smoking

Fitted Line Plot
smoking = 1.987 + 0.2837 income

S	1.67838
R-Sq	4.8%
R-Sq(adj)	4.3%

6.Conclusion

The usage of smoking in Kandy city, like in many urban areas, is influenced by social, cultural, and economic factors. Despite awareness campaigns, smoking remains common among some segments of the population, particularly in public spaces. However, government regulations and health concerns have led to increasing restrictions on smoking in public areas and growing awareness about its harmful effects. The overall trend is toward reducing smoking rates, but it continues to be a prevalent issue in Kandy.