1. The owner of a large chain of ice-cream stores, would like to predict the sales using atmospheric temperature as predictor. In order to study the relationship, he collected a data on the sale of 21 days randomly.

|  |  |  |
| --- | --- | --- |
| Day | Daily High  Temperature (0F) | Sales per store  (per hundred dollar) |
| 1 | 63 | 1.52 |
| 2 | 70 | 1.68 |
| 3 | 73 | 1.80 |
| 4 | 75 | 2.05 |
| 5 | 80 | 2.36 |
| 6 | 82 | 2.25 |
| 7 | 85 | 2.68 |
| 8 | 88 | 2.90 |
| 9 | 90 | 3.14 |
| 10 | 91 | 3.06 |
| 11 | 92 | 3.24 |
| 12 | 75 | 1.92 |
| 13 | 98 | 3.40 |
| 14 | 100 | 3.28 |
| 15 | 92 | 3.17 |
| 16 | 87 | 2.83 |
| 17 | 84 | 2.58 |
| 18 | 88 | 2.86 |
| 19 | 80 | 2.26 |
| 20 | 82 | 2.14 |
| 21 | 76 | 1.98 |

1. If you are a data scientist, and assigned a job to model this data using a simple linear regression, first draw a scatter diagram in order to check whether linear model is model is appropriate or not. Also report the type and strength of correlation between these variables.
2. Compute Karl Pearson’s correlation
3. Assuming linear relationship between these two variables develop a predicting equation using least square method.
4. Interpret the regression coefficients b0 and b1
5. Predict the average sales per store for a day in which temperature is 830 F
6. Compute coefficient of determination R2 and interpret the result
7. Compute Standard error of estimation and interpret the result
8. Carryout the residual analysis and make statements about model violations.
9. A quality characteristic of interest for a tea-bag filling process is the weight of the tea in the individual bags. If the bags are underfilled, two problems arise. First, customers may not be able to brew the tea to be as strong as they wish. Second the company may be in violation of the truth in labelling laws. The label weight in the packet is 5.5 gm. Temperature variation, humidity inside the factory, difference in the density of the tea, extremely fast filling process of machine might produce variation. A quality controller took a random sample of 50 tea-bags in an hour by a single machine. The data is shown below (data is already sorted in ascending order)

Weight of tea-bags in grams

5.25 5.29 5.32 5.32 5.34 5.36 5.40 5.40 5.40 5.41

5.42 5.42 5.44 5.44 5.44 5.45 5.45 5.46 5.47 5.47

5.49 5.50 5.50 5.50 5.51 5.52 5.53 5.53 5.53 5.53

5.54 5.54 5.55 5.55 5.56 5.56 5.57 5.57 5.57 5.58

5.58 5.58 5.61 5.61 5.62 5.63 5.65 5.67 5.67 5.77

Compute

1. Compute arithmetic mean
2. Compute median
3. Compute midrange
4. Compute range
5. Compute first quartile
6. Compute third quartile
7. Compute tenth percentile
8. Compute ninetieth percentile
9. Compute inter-quartile range
10. Compute variance and standard deviation
11. Compute coefficient of variation
12. Compute coefficient of skewness
13. Compute coefficient of kurtosis
14. Construct box-and-whisker plot
15. Make a summary of analysis, mentioning centrality, variation, shape and size of distribution.
16. Based on your analysis in (o) is the product is underfilled or overfilled or working well? Give reasons