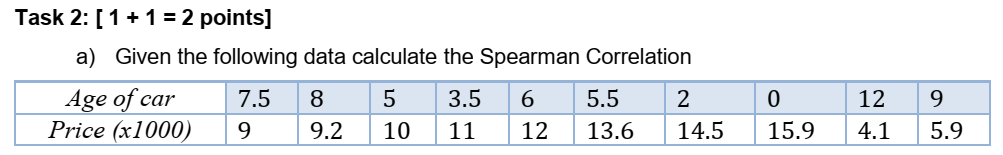
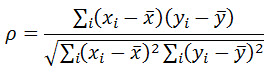


Solution Task-1:

|  |  |  |  |
| --- | --- | --- | --- |
| 0.721242 | 0.165522 | -0.88417 | 1.894429 |
| 0.536002 | -0.14321 | -1.06941 | 2.203163 |
| 0.227269 | 0.227269 | -0.94592 | 1.523949 |
| 0.597749 | -0.20496 | -0.94592 | 1.215216 |
| 0.289015 | 0.103775 | -1.00767 | 2.203163 |
| 0.659495 | -0.14321 | -0.51369 | 1.400456 |
| 0.227269 | -0.20496 | -1.00767 | 1.709189 |
| 0.289015 | 0.165522 | -0.94592 | 2.079669 |
| 0.412509 | -0.20496 | -0.88417 | 1.709189 |
| 0.721242 | 0.227269 | -1.00767 | 1.523949 |
| 0.968229 | -0.08146 | -1.06941 | 2.203163 |
| 0.350762 | -0.01972 | -0.94592 | -0.2667 |
| 0.906482 | 0.103775 | -1.00767 | 1.709189 |
| 0.782989 | 0.350762 | -1.06941 | 1.956176 |

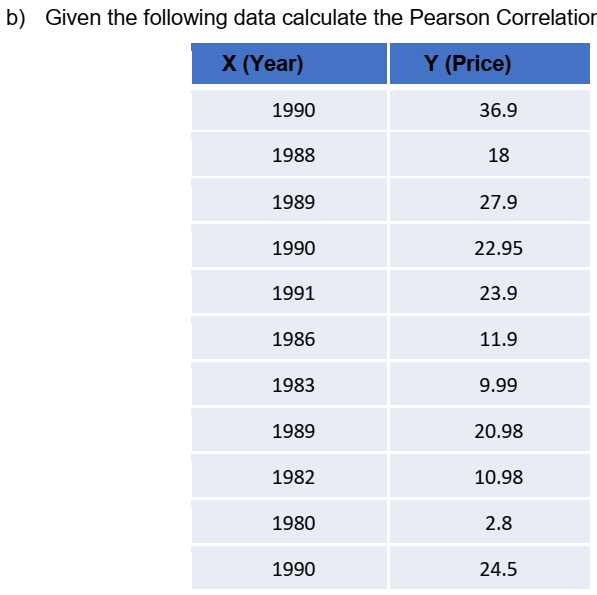
Outliers: No outlier is found in this data.

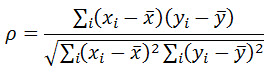




|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |  |
| Age | 7.5 | 8 | 5 | 3.5 | 6 | 5.5 | 2 | 0 | 12 | 9 |  |
| Price | 9 | 9.2 | 10 | 11 | 12 | 13.6 | 14.5 | 15.9 | 4.1 | 5.9 |  |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 7.5 | 8 | 5 | 3.5 | 6 | 5.5 | 2 | 0 | 12 | 9 | 5.85 |
| Price | 9 | 9.2 | 10 | 11 | 12 | 13.6 | 14.5 | 15.9 | 4.1 | 5.9 | 10.52 |
| (xi-x')\*(yi-y') | -2.508 | -2.838 | 0.442 | -1.128 | 0.222 | -1.078 | -15.323 | -31.473 | -39.483 | -14.553 | -107.72 |
| (xi-x')^2 | 2.7225 | 4.6225 | 0.7225 | 5.5225 | 0.0225 | 0.1225 | 14.8225 | 34.2225 | 37.8225 | 9.9225 | 110.525 |
| (yi-y')^2 | 55.9504 | 58.9824 | 71.9104 | 89.8704 | 109.8304 | 145.9264 | 168.4804 | 206.7844 | 6.6564 | 19.1844 | 933.576 |
|  |  |  |  |  |  |  |  |  |  |  | -0.335344543 |





**Task 3:**

Cross validation is the process of splitting data into two sets i.e. training set and test set. The statistical model is trained on training set whereas test set is used to find the predictions and then the difference between predictions and actual outcomes given in test set is used to measure the accuracy of results found by the model.

K fold process divides into user specified number of folds i.e. groups. If we are using 3 fold cross validation, the data is divided into three groups. In start first group is used for validation whereas other two groups are used for training. In next iteration, first and 3rd groups are used for training and 2nd group is used for testing. Finally first two groups are used for training and last group is used for testing as shown below:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

The shaded regions of our data are used for training and white part is used for validation.

Stratification is the process in which the algorithm attempts to make equal sized splits during k-fold cross validation.