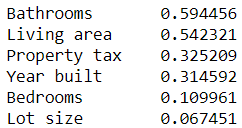
**Suggested Solution for Practice Paper**

1a) 

1b) Lot size since it has the highest p value = 0.831

1c)

F=13.34, p value = 2.42e-10

Reject 𝐻0.

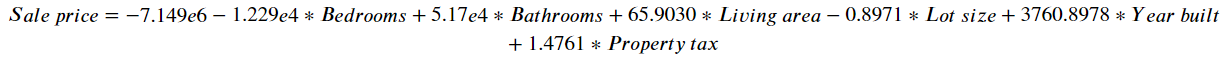
1d) Estimate of residual std

= RMSE

= (by adj Rsquare, by Rsquare)

= (47359.4566, 45679.6665)

1e)



1f) Predicted value

= (exact value, value calculated by hand)

= (265360.18921654, 277472.75100000086)

1g) R\_squared = 0.506  
∴ 50.6% of variation explained

2a) Coefficient standard errors increase/Fewer statistically significant slopes (t-ratios decrease and p-values increase)/Difficulty interpreting coefficients/Coefficients change as others come and go

2b) drop the x-variable from the regression/ combine it with other x-variable(s)

2c) test statistics= -0.10065443193208266

Since 0.10065443193208266 is less than 1.494, p value > 0.142, we can not reject H0

2d) R2 cannot decrease when another independent variable x is added to the regression/ Adjusted R2 gives penalty to the increase in numbers of predictors

2e) Yes, the model violates with the assumption of normally distributed residual since the QQ plot shows the residuals are not normally distributed/ skewness > 1/ Kurtosis >> 3.  
(independence assumption of fitted values and residuals.)

2f) 

2g) Predicted value

= (exact value, value calculated by hand)

= (158.28771429, 158.1784)

3ai) When ‘lstat’ increases 1 unit, we have 95% confidence that ‘medv’ decreases by at least 4.3% and at most 4.9%.

3aii) ;

3bi) I agree with him as the two models both have only 1 predictor and the responses of the two model are both ‘log(medv)’.

3bii) ;

3biii) I don’t agree with him. The residual plot shows a funnel shape which means the variance decreases when the fitted value increases. Natural log transformation is not necessary to correct the heteroskedasticity problem.

4a)



('Renter' can be replaced by 'Owner', 'Male' can be replaced by 'Female', 'M' or 'F'; order doesn't matter)

4b)



4c) Yes because the 95% C.I of males encloses negative numbers only. So the balance of male is significantly lower than female, given that other conditions remain.

4d)



The intercept 𝛽0 increases from 14.3475 to 22.125, and the slope of *Expense* 𝛽2 increases from -4.9187 to 7.7264.

5a) For size of 1, the selected model is {X1} since it has the largest R-square;

For size of 2, the selected model is {X1,X2} since it has the largest R-square;

For size of 3, the selected model is {X1,X3,X4} since it has the largest R-square;

For size of 4, the selected model is {X1,X2,X3,X4} since it has the largest R-square (or it is the only model with size 4.

So the candidate models are {X1}, {X1,X2}, {X1,X3,X4}, {X1,X2,X3,X4}.

5b) Since model{X1,X2,X3,X4} has the largest adjusted R-square in the candidate list, so it is the single best model

5c) 1+4(5)/2 = 11 (if we include the null model).

5d) For size of 1, the selected model is {X1} since it has the largest R-square;

For size of 2, the selected model is {X1,X2} since it has the largest R-square among {X1,X2}, {X1,X3} and {X1,X4};

For size of 3, the selected model is {X1,X2,X3} since it has the largest R-square between {X1,X2,X3} and {X1,X2,X4} ;

For size of 4, the selected model is {X1,X2,X3,X4} since it has the largest R-square (or it is the only model with size 4.

So the candidate models are {X1}, {X1,X2}, {X1,X2,X3}, {X1,X2,X3,X4}.

6a)

The predicted Y and square error for ith 5-fold CV is summarized as below

|  |  |  |
| --- | --- | --- |
| Y | Predict Y | Square Error |
| 271 | 119.82 | 22855.39 |
| 152 | 191.35 | 1548.42 |
| 274 | 58.87 | 46280.92 |
| 183 | 250.98 | 4621.28 |
| 135 | 26.99 | 11666.16 |
|  | LOOCV | 17394.43 |

So the 5-fold CV estimate is 17394.43.

6b)

The predicted Y and square error for ith 5-fold CV is summarized as below

|  |  |  |
| --- | --- | --- |
| Y | Predict Y | Square Error |
| 271 | 179.11 | 8443.77 |
| 152 | 227.59 | 5713.85 |
| 274 | 154.9 | 14184.81 |
| 183 | 244.19 | 3744.22 |
| 135 | 897.05 | 580720.2 |
|  | LOOCV | 122561.37 |

So the 5-fold CV estimate is 122561.37

6c) As the 5-fold CV estimate of Model A is smaller, so Model A is better.

7a) Maximum Likelihood Estimation.

7b) log (p/(1-p)) = -15.4255+0.0046(duration)-0.0064(nr\_employed)+1.9147(poutcome\_success)-0.4837(emp\_var\_rate)+0.0767(previous)+0.5958(poutcome\_nonexistent)-0.3073(contact\_telephone)+1.3251(month\_mar)+0.1288(month\_oct)+0.4652(cons\_price\_idx)-0.1861(month\_sep)-0.8757(month\_may)+0.3495(default\_no)+ 0.4418(job\_student)+0.3845(job\_retired)

7c) The odds ratio will increase by (100\*0.4652)%= 46.52% per unit increase in ‘cons\_price\_idx’.

7d) 0.086845929

8a) Accuracy: (26946+1844) / (26946+1844+1911+2249) = 0.873748

8b) FPR: 2249 / (26946+2249) = 0.077033

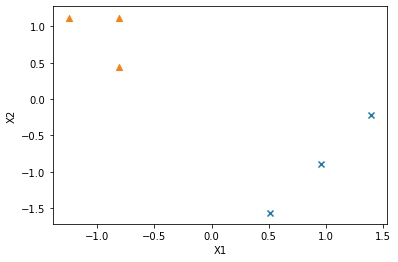
TPR: 1844 / (1844 + 1911) = 0.491079

8c) The data is balanced that TP + FN = FP + TN

9a)

|  |  |  |
| --- | --- | --- |
|  | X1 | X2 |
| 0 |  |  |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

9b)

9c)

9d) No, we can see that the plot does not decrease dramatically after k=4.

10a) Distance matrix =

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Index | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 |  |  |  |  |  |  |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |

where the cell represents the distance between th observation and th observation.

The minimum distance is that between index 0 and index 1. Therefore, group index 0 and index 1 into a new cluster (0, 1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Index | (0, 1) | 2 | 3 | 4 | 5 |
| (0, 1) |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |

The distance between cluster (0, 1) and 2 is the minimum distance between distance between index 0 and index 2 and that between index 1 and index 2 as we use single linkage.

The minimum distance is that between index 3 and index 5. Therefore, group index 3 and index 5 into a new cluster (3, 5).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | (0, 1) | 2 | (3, 5) | 4 |
| (0, 1) |  |  |  |  |
| 2 |  |  |  |  |
| (3, 5) |  |  |  |  |
| 4 |  |  |  |  |

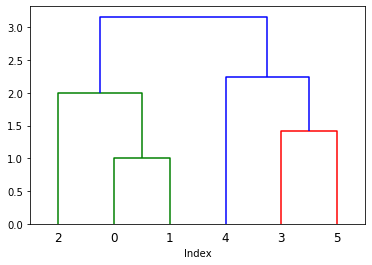
The minimum distance is that between cluster (0, 1) and index 2. Therefore, group cluster (0, 1) and index 2 into a new cluster ((0, 1), 2).

|  |  |  |  |
| --- | --- | --- | --- |
| Index | ((0, 1), 2) | (3, 5) | 4 |
| ((0, 1), 2) |  |  |  |
| (3, 5) |  |  |  |
| 4 |  |  |  |

The minimum distance is that between cluster (3, 5) and index 4. Therefore, group cluster (3, 5) and index 4 into a new cluster ((3, 5), 4).

|  |  |  |
| --- | --- | --- |
| Index | ((0, 1), 2) | ((3, 5), 4) |
| ((0, 1), 2) |  |  |
| ((3, 5), 4) |  |  |

The minimum distance is that between cluster ((0, 1), 2) and cluster ((3, 5), 4). Therefore, group cluster ((0, 1), 2) and cluster ((3, 5), 4) into a new cluster (((0, 1), 2), ((3, 5), 4)).



10b) When number of clusters is 3, the three clusters are ((0, 1), 2), (3, 5), 4.

10c) No, the hierarchical clustering labels the data by the structure. A new data set could have a completely different structure from the discovered one. So we would rerun the whole algorithm to create a new set of labels.