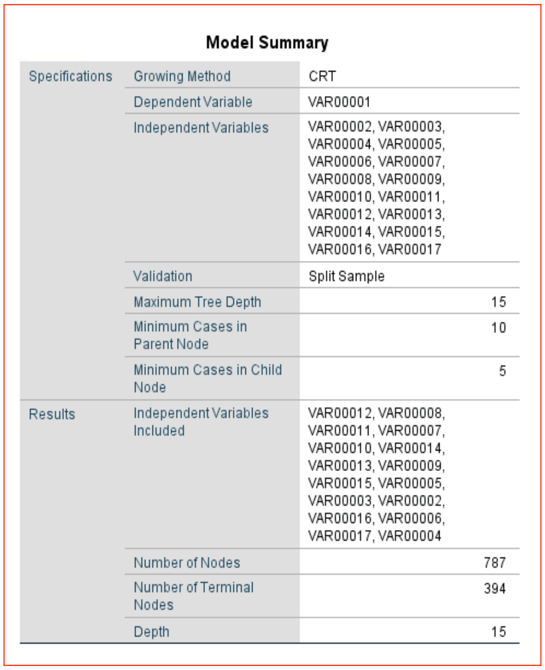
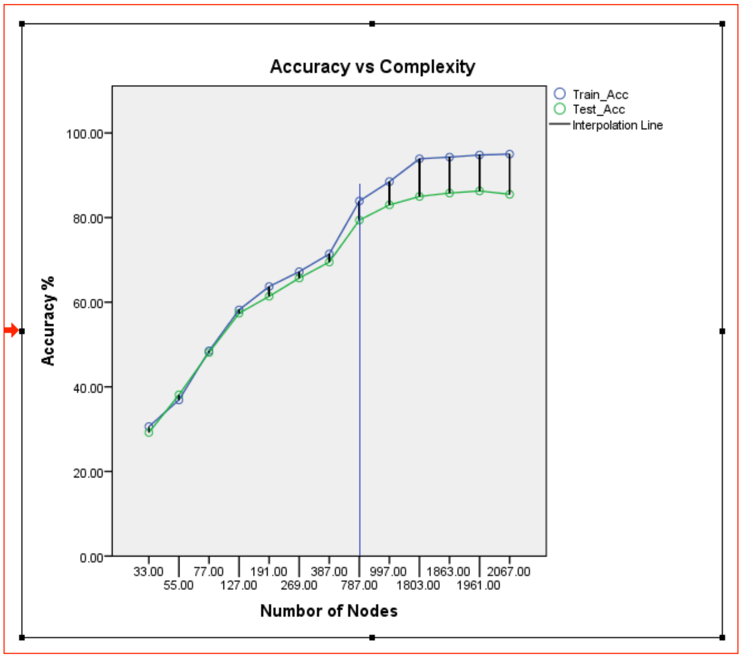
Question 1-a
Ans: The following shows the Accuracy of the Decision Tree model:
Variables:
VAR00001 - capital letter (26 values from A to Z) ------Independent / Target Variable
VAR00002 - x-box horizontal position of box (integer)
VAR00003 - y-box vertical position of box (integer)
VAR00004 - width width of box (integer)
VAR00005 - high height of box (integer)
VAR00006 - onpix total # on pixels (integer)
VAR00007 - x-bar mean x of on pixels in box (integer)
VAR00008 - y-bar mean y of on pixels in box (integer)
VAR00009 - x2bar mean x variance (integer)
VAR00010 - y2bar mean y variance (integer)
VAR00011 - xybar mean x y correlation (integer)
VAR00012 - x2ybr mean of x * x * y (integer)
VAR00013 - xy2br mean of x * y * y (integer)
VAR00014 - x-ege mean edge count left to right (integer)
VAR00015 - xegvy correlation of x-ege with y (integer)
VAR00016 - y-ege mean edge count bottom to top (integer)
VAR00017 - yegvx correlation of y-ege with x (integer)



**Model Summary**

| | | |
|---|---|---|
| Specifications | Growing Method | CRT |
| | Dependent Variable | VAR00001 |
| | Independent Variables | VAR00002, VAR00003, VAR00004, VAR00005, VAR00006, VAR00007, VAR00008, VAR00009, VAR00010, VAR00011, VAR00012, VAR00013, VAR00014, VAR00015, VAR00016, VAR00017 |
| | Validation | Split Sample |
| | Maximum Tree Depth | 15 |
| | Minimum Cases in Parent Node | 10 |
| | Minimum Cases in Child Node | 5 |
| Results | Independent Variables Included | VAR00012, VAR00008, VAR00011, VAR00007, VAR00010, VAR00014, VAR00013, VAR00009, VAR00015, VAR00005, VAR00003, VAR00002, VAR00016, VAR00006, VAR00017, VAR00004 |
| | Number of Nodes | 787 |
| | Number of Terminal Nodes | 394 |
| | Depth | 15 |

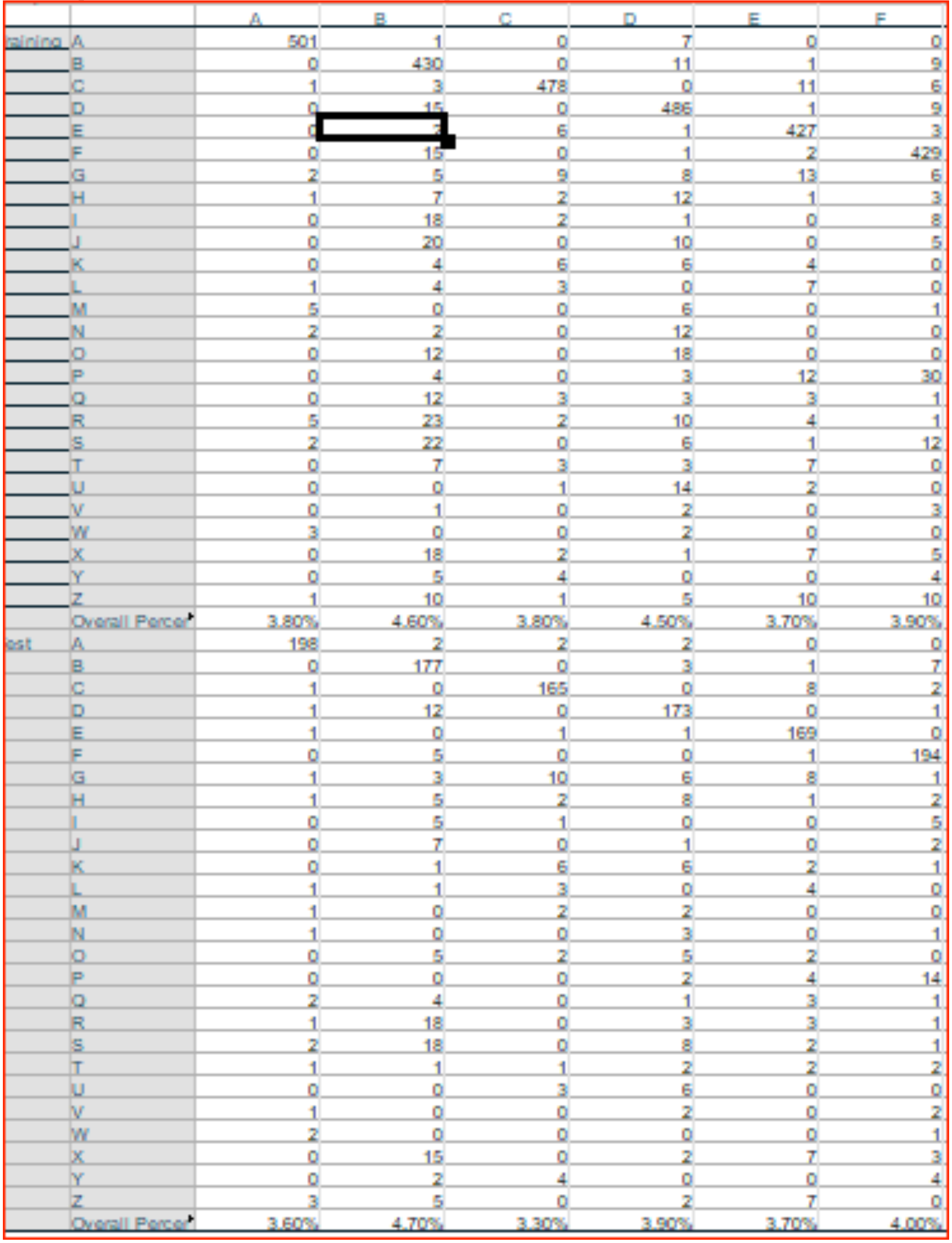


By observing the above Accuracy plot, the best model I would propose for this particular dataset including the following parameters, minimum of cases for parents is 10 and for child is 5, the stopping condition is with depth 15. In this model, the number of nodes is 797 with 394 terminal nodes. The sample split validation setting is 70% Training and 30% Testing data. The above result implies another feature that SPSS algorithm does push the tree to the maximum depth as preset.

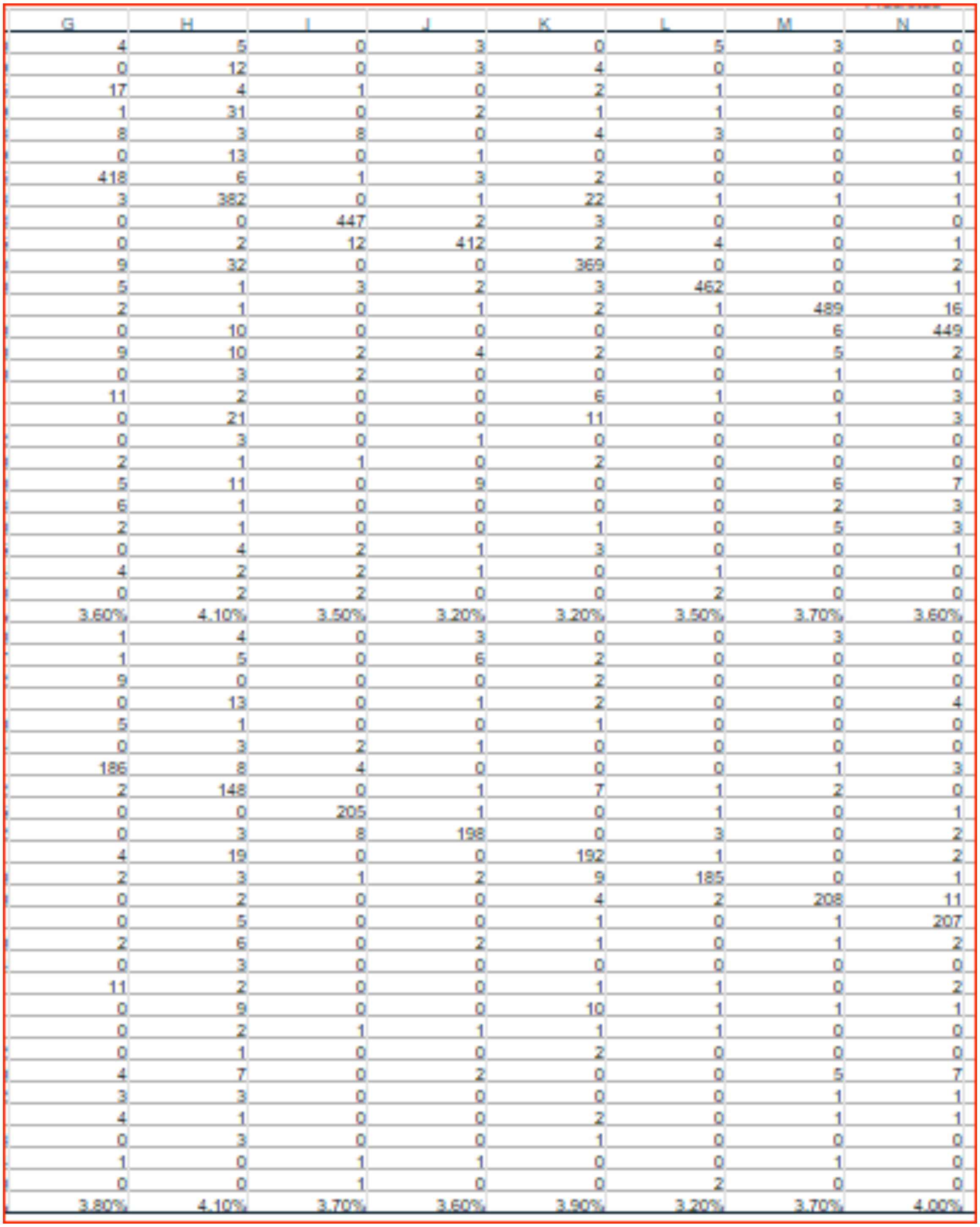
IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

Question 1-b: Report the misclassification matrix and interpret



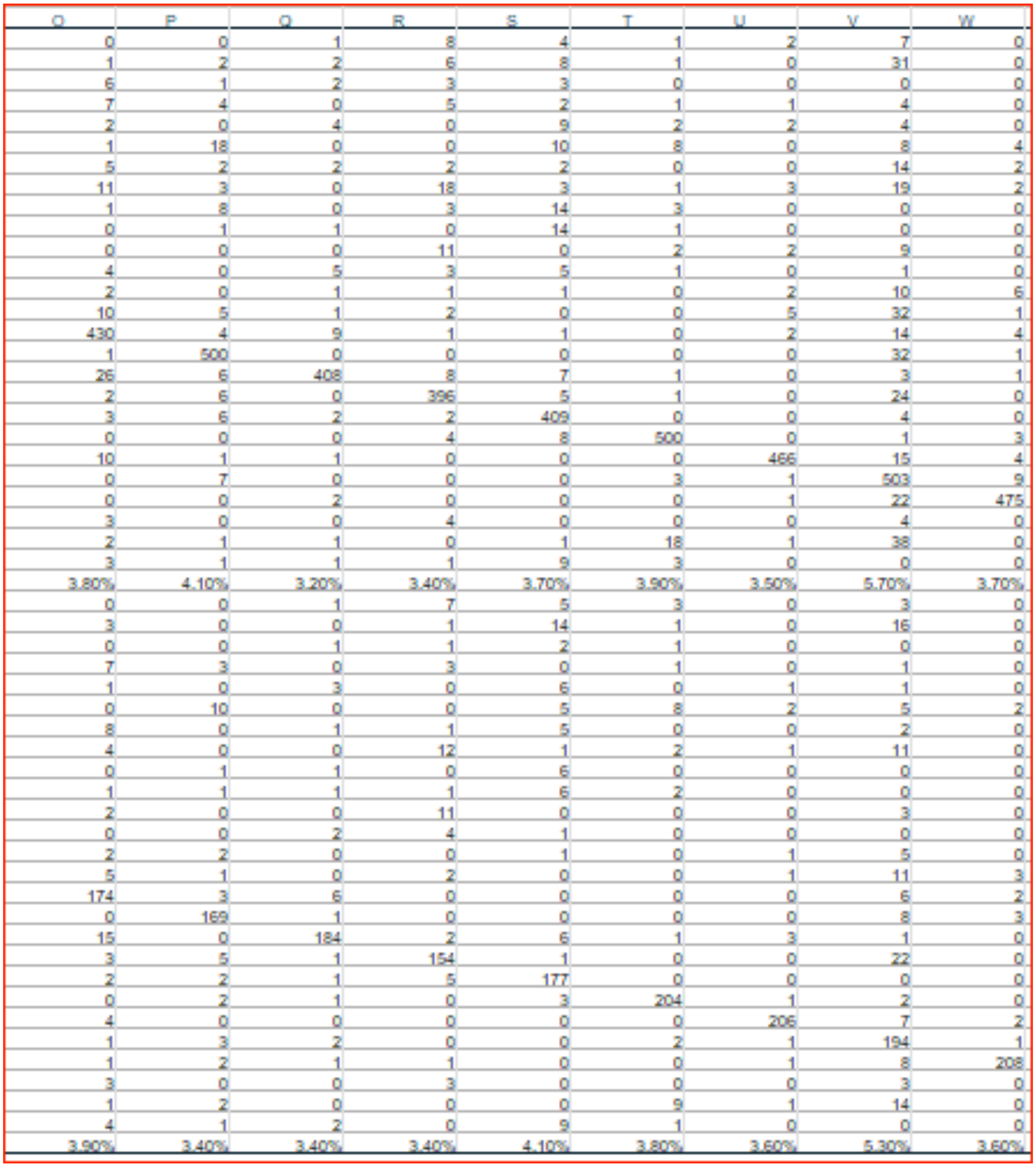
| | | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|
| raining | A | 501 | 1 | 0 | 7 | 0 | 0 |
| | B | 0 | 430 | 0 | 11 | 1 | 9 |
| | C | 1 | 3 | 478 | 0 | 11 | 6 |
| | D | 0 | 15 | 0 | 486 | 1 | 9 |
| | E | 0 | 2 | 6 | 1 | 427 | 3 |
| | F | 0 | 15 | 0 | 1 | 2 | 429 |
| | G | 2 | 5 | 9 | 8 | 13 | 6 |
| | H | 1 | 7 | 2 | 12 | 1 | 3 |
| | I | 0 | 18 | 2 | 1 | 0 | 8 |
| | J | 0 | 20 | 0 | 10 | 0 | 5 |
| | K | 0 | 4 | 6 | 6 | 4 | 0 |
| | L | 1 | 4 | 3 | 0 | 7 | 0 |
| | M | 5 | 0 | 0 | 6 | 0 | 1 |
| | N | 2 | 2 | 0 | 12 | 0 | 0 |
| | O | 0 | 12 | 0 | 18 | 0 | 0 |
| | P | 0 | 4 | 0 | 3 | 12 | 30 |
| | Q | 0 | 12 | 3 | 3 | 3 | 1 |
| | R | 5 | 23 | 2 | 10 | 4 | 1 |
| | S | 2 | 22 | 0 | 6 | 1 | 12 |
| | T | 0 | 7 | 3 | 3 | 7 | 0 |
| | U | 0 | 0 | 1 | 14 | 2 | 0 |
| | V | 0 | 1 | 0 | 2 | 0 | 3 |
| | W | 3 | 0 | 0 | 2 | 0 | 0 |
| | X | 0 | 18 | 2 | 1 | 7 | 5 |
| | Y | 0 | 5 | 4 | 0 | 0 | 4 |
| | Z | 1 | 10 | 1 | 5 | 10 | 10 |
| | Overall Percen | 3.80% | 4.60% | 3.80% | 4.50% | 3.70% | 3.90% |
| est | A | 198 | 2 | 2 | 2 | 0 | 0 |
| | B | 0 | 177 | 0 | 3 | 1 | 7 |
| | C | 1 | 0 | 165 | 0 | 8 | 2 |
| | D | 1 | 12 | 0 | 173 | 0 | 1 |
| | E | 1 | 0 | 1 | 1 | 169 | 0 |
| | F | 0 | 5 | 0 | 0 | 1 | 194 |
| | G | 1 | 3 | 10 | 6 | 8 | 1 |
| | H | 1 | 5 | 2 | 8 | 1 | 2 |
| | I | 0 | 5 | 1 | 0 | 0 | 5 |
| | J | 0 | 7 | 0 | 1 | 0 | 2 |
| | K | 0 | 1 | 6 | 6 | 2 | 1 |
| | L | 1 | 1 | 3 | 0 | 4 | 0 |
| | M | 1 | 0 | 2 | 2 | 0 | 0 |
| | N | 1 | 0 | 0 | 3 | 0 | 1 |
| | O | 0 | 5 | 2 | 5 | 2 | 0 |
| | P | 0 | 0 | 0 | 2 | 4 | 14 |
| | Q | 2 | 4 | 0 | 1 | 3 | 1 |
| | R | 1 | 18 | 0 | 3 | 3 | 1 |
| | S | 2 | 18 | 0 | 8 | 2 | 1 |
| | T | 1 | 1 | 1 | 2 | 2 | 2 |
| | U | 0 | 0 | 3 | 6 | 0 | 0 |
| | V | 1 | 0 | 0 | 2 | 0 | 2 |
| | W | 2 | 0 | 0 | 0 | 0 | 1 |
| | X | 0 | 15 | 0 | 2 | 7 | 3 |
| | Y | 0 | 2 | 4 | 0 | 0 | 4 |
| | Z | 3 | 5 | 0 | 2 | 7 | 0 |
| | Overall Percen | 3.60% | 4.70% | 3.30% | 3.90% | 3.70% | 4.00% |



| G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|
| 4 | 5 | 0 | 3 | 0 | 5 | 3 | 0 |
| 0 | 12 | 0 | 3 | 4 | 0 | 0 | 0 |
| 17 | 4 | 1 | 0 | 2 | 1 | 0 | 0 |
| 1 | 31 | 0 | 2 | 1 | 1 | 0 | 6 |
| 8 | 3 | 8 | 0 | 4 | 3 | 0 | 0 |
| 0 | 13 | 0 | 1 | 0 | 0 | 0 | 0 |
| 418 | 6 | 1 | 3 | 2 | 0 | 0 | 1 |
| 3 | 382 | 0 | 1 | 22 | 1 | 1 | 1 |
| 0 | 0 | 447 | 2 | 3 | 0 | 0 | 0 |
| 0 | 2 | 12 | 412 | 2 | 4 | 0 | 1 |
| 9 | 32 | 0 | 0 | 369 | 0 | 0 | 2 |
| 5 | 1 | 3 | 2 | 3 | 462 | 0 | 1 |
| 2 | 1 | 0 | 1 | 2 | 1 | 489 | 16 |
| 0 | 10 | 0 | 0 | 0 | 0 | 6 | 449 |
| 9 | 10 | 2 | 4 | 2 | 0 | 5 | 2 |
| 0 | 3 | 2 | 0 | 0 | 0 | 1 | 0 |
| 11 | 2 | 0 | 0 | 6 | 1 | 0 | 3 |
| 0 | 21 | 0 | 0 | 11 | 0 | 1 | 3 |
| 0 | 3 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 0 | 2 | 0 | 0 | 0 |
| 5 | 11 | 0 | 9 | 0 | 0 | 6 | 7 |
| 6 | 1 | 0 | 0 | 0 | 0 | 2 | 3 |
| 2 | 1 | 0 | 0 | 1 | 0 | 5 | 3 |
| 0 | 4 | 2 | 1 | 3 | 0 | 0 | 1 |
| 4 | 2 | 2 | 1 | 0 | 1 | 0 | 0 |
| 0 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 3.60% | 4.10% | 3.50% | 3.20% | 3.20% | 3.50% | 3.70% | 3.60% |
| 1 | 4 | 0 | 3 | 0 | 0 | 3 | 0 |
| 1 | 5 | 0 | 6 | 2 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 0 | 13 | 0 | 1 | 2 | 0 | 0 | 4 |
| 5 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 3 | 2 | 1 | 0 | 0 | 0 | 0 |
| 186 | 8 | 4 | 0 | 0 | 0 | 1 | 3 |
| 2 | 148 | 0 | 1 | 7 | 1 | 2 | 0 |
| 0 | 0 | 205 | 1 | 0 | 1 | 0 | 1 |
| 0 | 3 | 8 | 198 | 0 | 3 | 0 | 2 |
| 4 | 19 | 0 | 0 | 192 | 1 | 0 | 2 |
| 2 | 3 | 1 | 2 | 9 | 185 | 0 | 1 |
| 0 | 2 | 0 | 0 | 4 | 2 | 208 | 11 |
| 0 | 5 | 0 | 0 | 1 | 0 | 1 | 207 |
| 2 | 6 | 0 | 2 | 1 | 0 | 1 | 2 |
| 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 2 | 0 | 0 | 1 | 1 | 0 | 2 |
| 0 | 9 | 0 | 0 | 10 | 1 | 1 | 1 |
| 0 | 2 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 | 7 | 0 | 2 | 0 | 0 | 5 | 7 |
| 3 | 3 | 0 | 0 | 0 | 0 | 1 | 1 |
| 4 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 |
| 3.80% | 4.10% | 3.70% | 3.60% | 3.90% | 3.20% | 3.70% | 4.00% |

IS467<br>
Homework#4<br>
Date: 06-03-2017

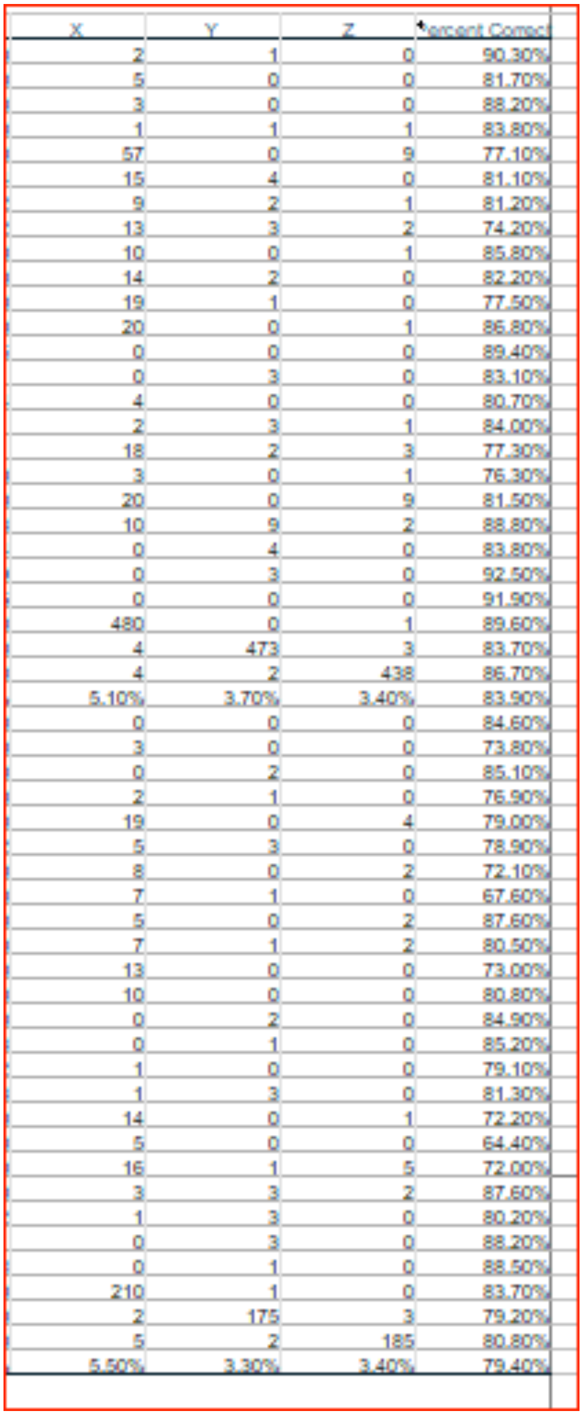
Name: Kai Chung, Ying<br>
Email: kying@mail.depaul.edu



| O | P | Q | R | S | T | U | V | W |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 8 | 4 | 1 | 2 | 7 | 0 |
| 1 | 2 | 2 | 6 | 8 | 1 | 0 | 31 | 0 |
| 6 | 1 | 2 | 3 | 3 | 0 | 0 | 0 | 0 |
| 7 | 4 | 0 | 5 | 2 | 1 | 1 | 4 | 0 |
| 2 | 0 | 4 | 0 | 9 | 2 | 2 | 4 | 0 |
| 1 | 18 | 0 | 0 | 10 | 8 | 0 | 8 | 4 |
| 5 | 2 | 2 | 2 | 2 | 0 | 0 | 14 | 2 |
| 11 | 3 | 0 | 18 | 3 | 1 | 3 | 19 | 2 |
| 1 | 8 | 0 | 3 | 14 | 3 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 14 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 11 | 0 | 2 | 2 | 9 | 0 |
| 4 | 0 | 5 | 3 | 5 | 1 | 0 | 1 | 0 |
| 2 | 0 | 1 | 1 | 1 | 0 | 2 | 10 | 6 |
| 10 | 5 | 1 | 2 | 0 | 0 | 5 | 32 | 1 |
| 430 | 4 | 9 | 1 | 1 | 0 | 2 | 14 | 4 |
| 1 | 500 | 0 | 0 | 0 | 0 | 0 | 32 | 1 |
| 26 | 6 | 408 | 8 | 7 | 1 | 0 | 3 | 1 |
| 2 | 6 | 0 | 396 | 5 | 1 | 0 | 24 | 0 |
| 3 | 6 | 2 | 2 | 409 | 0 | 0 | 4 | 0 |
| 0 | 0 | 0 | 4 | 8 | 500 | 0 | 1 | 3 |
| 10 | 1 | 1 | 0 | 0 | 0 | 466 | 15 | 4 |
| 0 | 7 | 0 | 0 | 0 | 3 | 1 | 503 | 9 |
| 0 | 0 | 2 | 0 | 0 | 0 | 1 | 22 | 475 |
| 3 | 0 | 0 | 4 | 0 | 0 | 0 | 4 | 0 |
| 2 | 1 | 1 | 0 | 1 | 18 | 1 | 38 | 0 |
| 3 | 1 | 1 | 1 | 9 | 3 | 0 | 0 | 0 |
| 3.80% | 4.10% | 3.20% | 3.40% | 3.70% | 3.90% | 3.50% | 5.70% | 3.70% |
| 0 | 0 | 1 | 7 | 5 | 3 | 0 | 3 | 0 |
| 3 | 0 | 0 | 1 | 14 | 1 | 0 | 16 | 0 |
| 0 | 0 | 1 | 1 | 2 | 1 | 0 | 0 | 0 |
| 7 | 3 | 0 | 3 | 0 | 1 | 0 | 1 | 0 |
| 1 | 0 | 3 | 0 | 6 | 0 | 1 | 1 | 0 |
| 0 | 10 | 0 | 0 | 5 | 8 | 2 | 5 | 2 |
| 8 | 0 | 1 | 1 | 5 | 0 | 0 | 2 | 0 |
| 4 | 0 | 0 | 12 | 1 | 2 | 1 | 11 | 0 |
| 0 | 1 | 1 | 0 | 6 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 6 | 2 | 0 | 0 | 0 |
| 2 | 0 | 0 | 11 | 0 | 0 | 0 | 3 | 0 |
| 0 | 0 | 2 | 4 | 1 | 0 | 0 | 0 | 0 |
| 2 | 2 | 0 | 0 | 1 | 0 | 1 | 5 | 0 |
| 5 | 1 | 0 | 2 | 0 | 0 | 1 | 11 | 3 |
| 174 | 3 | 6 | 0 | 0 | 0 | 0 | 6 | 2 |
| 0 | 169 | 1 | 0 | 0 | 0 | 0 | 8 | 3 |
| 15 | 0 | 184 | 2 | 6 | 1 | 3 | 1 | 0 |
| 3 | 5 | 1 | 154 | 1 | 0 | 0 | 22 | 0 |
| 2 | 2 | 1 | 5 | 177 | 0 | 0 | 0 | 0 |
| 0 | 2 | 1 | 0 | 3 | 204 | 1 | 2 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 206 | 7 | 2 |
| 1 | 3 | 2 | 0 | 0 | 2 | 1 | 194 | 1 |
| 1 | 2 | 1 | 1 | 0 | 0 | 1 | 8 | 208 |
| 3 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 0 |
| 1 | 2 | 0 | 0 | 0 | 9 | 1 | 14 | 0 |
| 4 | 1 | 2 | 0 | 9 | 1 | 0 | 0 | 0 |
| 3.90% | 3.40% | 3.40% | 3.40% | 4.10% | 3.80% | 3.60% | 5.30% | 3.60% |

IS467
Homework#4
Date: 06-03-2017

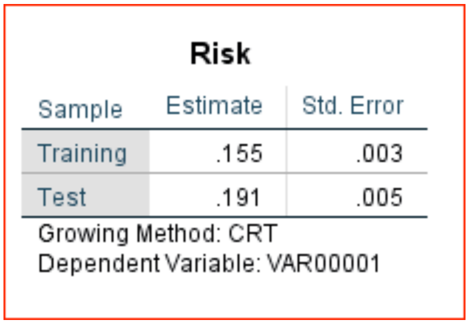
Name: Kai Chung, Ying
Email: kying@mail.depaul.edu



| X | Y | Z | Percent Correct |
|---|---|---|---|
| 2 | 1 | 0 | 90.30% |
| 5 | 0 | 0 | 81.70% |
| 3 | 0 | 0 | 88.20% |
| 1 | 1 | 1 | 83.80% |
| 57 | 0 | 9 | 77.10% |
| 15 | 4 | 0 | 81.10% |
| 9 | 2 | 1 | 81.20% |
| 13 | 3 | 2 | 74.20% |
| 10 | 0 | 1 | 85.80% |
| 14 | 2 | 0 | 82.20% |
| 19 | 1 | 0 | 77.50% |
| 20 | 0 | 1 | 86.80% |
| 0 | 0 | 0 | 89.40% |
| 0 | 3 | 0 | 83.10% |
| 4 | 0 | 0 | 80.70% |
| 2 | 3 | 1 | 84.00% |
| 18 | 2 | 3 | 77.30% |
| 3 | 0 | 1 | 76.30% |
| 20 | 0 | 9 | 81.50% |
| 10 | 9 | 2 | 88.80% |
| 0 | 4 | 0 | 83.80% |
| 0 | 3 | 0 | 92.50% |
| 0 | 0 | 0 | 91.90% |
| 480 | 0 | 1 | 89.60% |
| 4 | 473 | 3 | 83.70% |
| 4 | 2 | 438 | 86.70% |
| 5.10% | 3.70% | 3.40% | 83.90% |
| 0 | 0 | 0 | 84.60% |
| 3 | 0 | 0 | 73.80% |
| 0 | 2 | 0 | 85.10% |
| 2 | 1 | 0 | 76.90% |
| 19 | 0 | 4 | 79.00% |
| 5 | 3 | 0 | 78.90% |
| 8 | 0 | 2 | 72.10% |
| 7 | 1 | 0 | 67.60% |
| 5 | 0 | 2 | 87.60% |
| 7 | 1 | 2 | 80.50% |
| 13 | 0 | 0 | 73.00% |
| 10 | 0 | 0 | 80.80% |
| 0 | 2 | 0 | 84.90% |
| 0 | 1 | 0 | 85.20% |
| 1 | 0 | 0 | 79.10% |
| 1 | 3 | 0 | 81.30% |
| 14 | 0 | 1 | 72.20% |
| 5 | 0 | 0 | 64.40% |
| 16 | 1 | 5 | 72.00% |
| 3 | 3 | 2 | 87.60% |
| 1 | 3 | 0 | 80.20% |
| 0 | 3 | 0 | 88.20% |
| 0 | 1 | 0 | 88.50% |
| 210 | 1 | 0 | 83.70% |
| 2 | 175 | 3 | 79.20% |
| 5 | 2 | 185 | 80.80% |
| 5.50% | 3.30% | 3.40% | 79.40% |

The reason I have picked this model because the accuracy of the Testing data ( Blue Line) is 79.4% reaches the highest point of the curve with the least complexity (=787nodes). Also, the test set accuracy on the misclassication matrix are also acceptable range; the training and testing set accuracy ranges are very closed and the max percent accuracy of predicted value 1 reaches 88.5%. In addition, the standard deviation the 10-fold standard deviation (Shown below) is ONLY 0.005 with Estimate value 0.191 of the test set which is small enough to conclude that the model will not be overfitting. Also, this could imply that the Training and Testing data is split into an appropriate portion

(Training=70%/Testing=30%) for the classification model.



**Risk**

| Sample | Estimate | Std. Error |
|---|---|---|
| Training | .155 | .003 |
| Test | .191 | .005 |

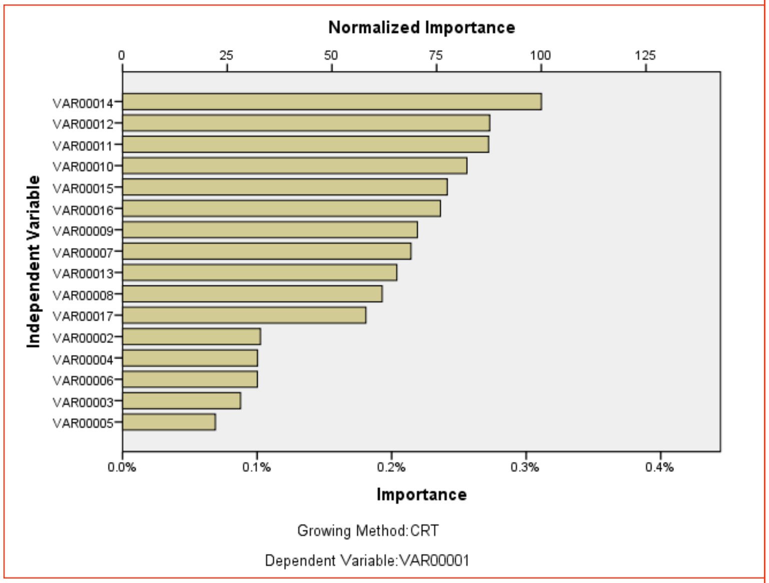
Growing Method: CRT
Dependent Variable: VAR00001

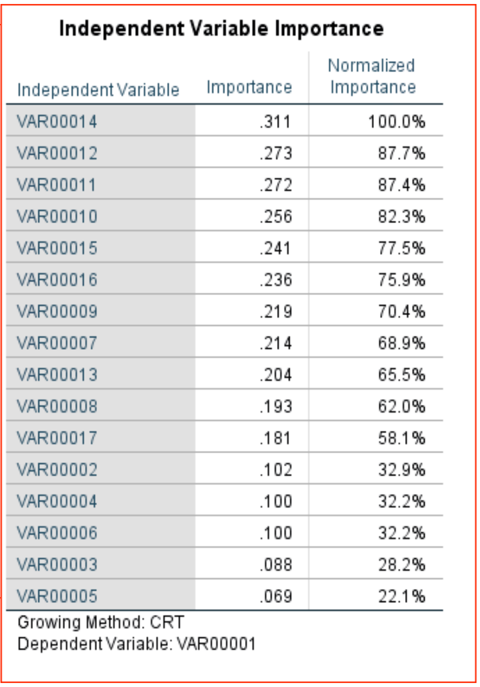
During the process of building the tree, I could observe that increasing the number of cases  allowed in parent and child nodes is decreasing the complexity (number of nodes) of the tree. The mechanism of this effect because increasing the number of cases of nodes would let the node NOT splitting until it reaches the minimum of cases we set. So that the higher values we set on the parent and child nodes, the less complexity of tree we are supposed to get. Also, we could observe the above plot, the least of the complexity we have (e.g. Number of nodes =33), the less accuracy we could get.

IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

Question 1-c: The most importance 3 attributes.





**Independent Variable Importance**

| Independent Variable | Importance | Normalized Importance |
|---|---|---|
| VAR00014 | .311 | 100.0% |
| VAR00012 | .273 | 87.7% |
| VAR00011 | .272 | 87.4% |
| VAR00010 | .256 | 82.3% |
| VAR00015 | .241 | 77.5% |
| VAR00016 | .236 | 75.9% |
| VAR00009 | .219 | 70.4% |
| VAR00007 | .214 | 68.9% |
| VAR00013 | .204 | 65.5% |
| VAR00008 | .193 | 62.0% |
| VAR00017 | .181 | 58.1% |
| VAR00002 | .102 | 32.9% |
| VAR00004 | .100 | 32.2% |
| VAR00006 | .100 | 32.2% |
| VAR00003 | .088 | 28.2% |
| VAR00005 | .069 | 22.1% |

Growing Method: CRT
Dependent Variable: VAR00001

Based on the above, the most important three attributes are shown as the following in order:

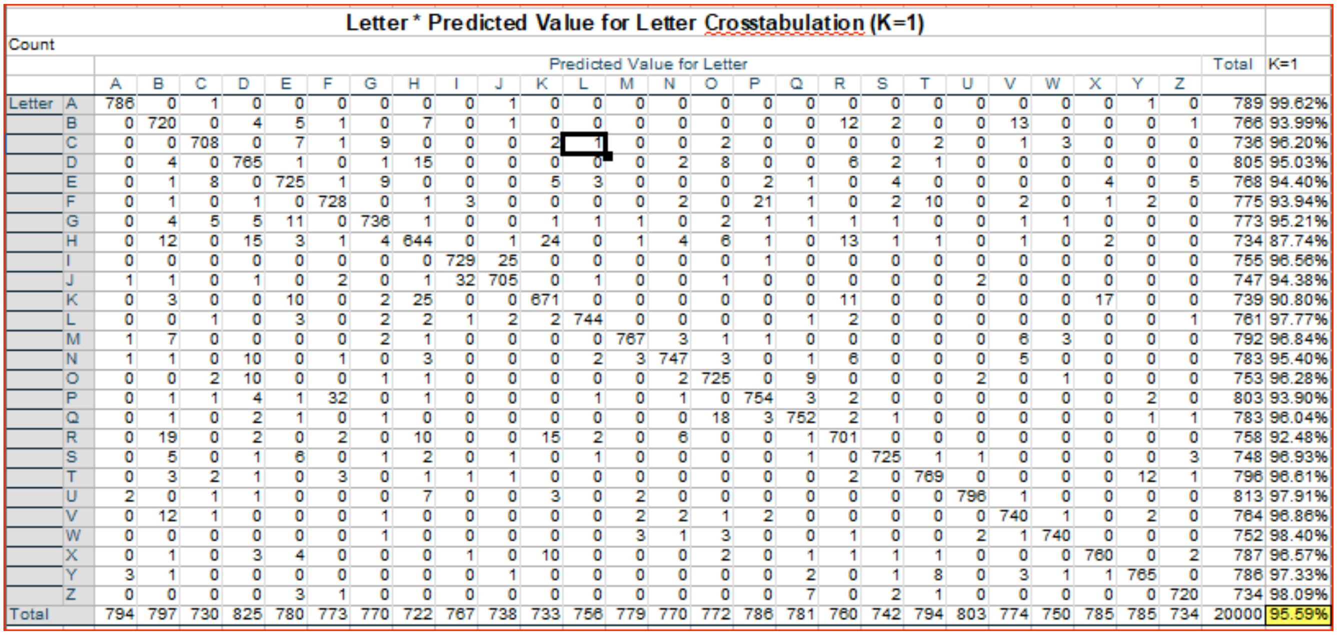
1) VAR00014 - x-ege mean edge count left to right (integer) ---100% (Normalized Importance)
2) VAR00012 - x2ybr mean of x * x * y (integer) ----------------87.7% (Normalized Importance)
3) VAR00011 - xybar mean x y correlation (integer) -------------87.4% (Normalized Importance)

Question 2-a:
Ans:
For this particular dataset, I have NOT transformed the Target variable to numeric number because they are single letters A-Z, there would be less chance to have errors in this case. On the other hand, the independent variables were transformed into smaller bin size(0-4=1,4-8=2,8-12=3,12-15=4) from the provided numerical integers. The reason to made this transformation is that 1.) reduce the algorithm running time 2.) To reduce the errors chances by the recorded numerical integers.

After running binned dataset, I tried to use the original dataset to run kNN again with the K (K=1,3,5,7). I found that the came up results are even more accurate. The partition ratio of the dataset is 70%Train / 30% Test data. The average accuracy of the original data result is ranged 94-96% and accuracy of the transformed dataset is ranged 78-80%. The detail matrix result is shown on part2-b. The reason might be the original data range are very standardized (1-15) for all the attributes. So that, the additional transformation / binning is not necessary for this case.

IS467 Name: Kai Chung, Ying
Homework#4 Email: kying@mail.depaul.edu
Date: 06-03-2017

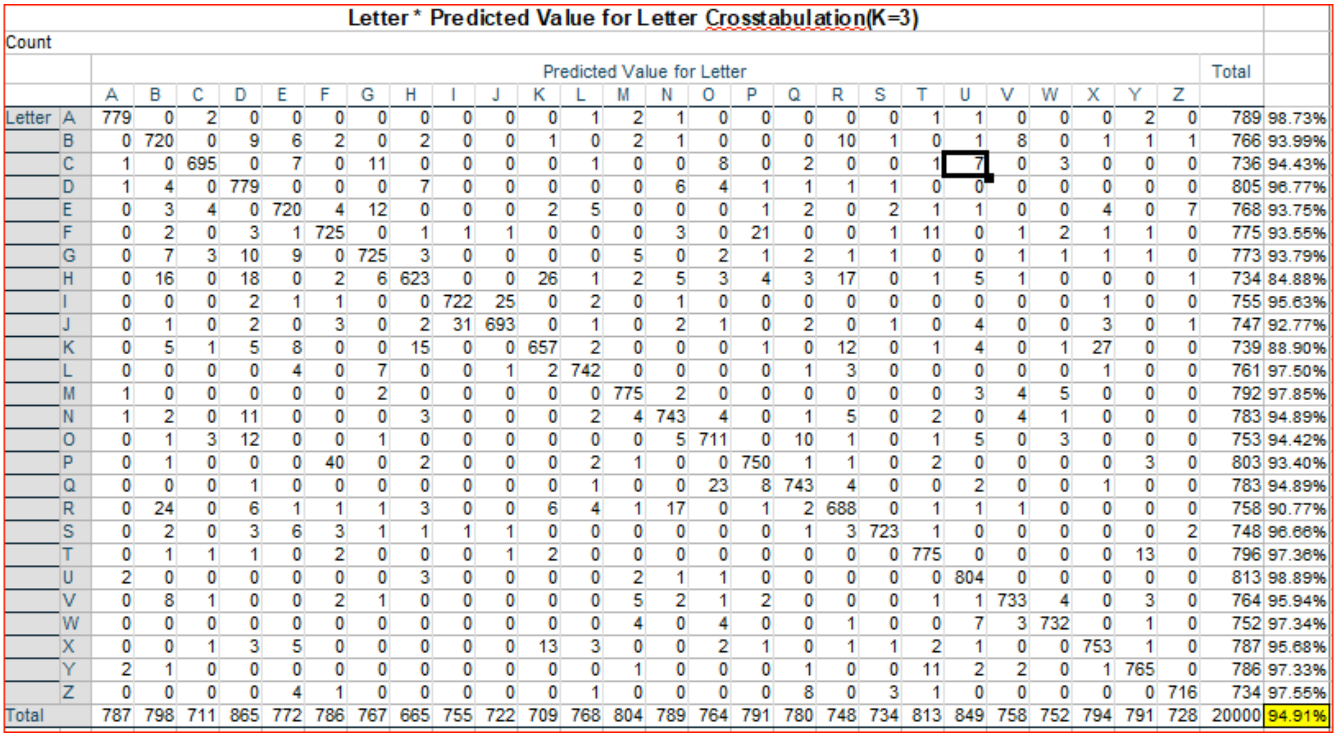
Question 2-b:
For the below shown matrix. The rows represent the provided labels count. The columns represent the predicted values count. The last column is the accuracy percentage of



**Letter * Predicted Value for Letter Crosstabulation (K=1)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | K=1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 786 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 789 | 99.62% |
| | B | 0 | 720 | 0 | 4 | 5 | 1 | 0 | 7 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 2 | 0 | 0 | 13 | 0 | 0 | 0 | 1 | 766 | 93.99% |
| | C | 0 | 0 | 708 | 0 | 7 | 1 | 9 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 3 | 0 | 0 | 0 | 736 | 96.20% |
| | D | 0 | 4 | 0 | 765 | 1 | 0 | 1 | 15 | 0 | 0 | 0 | 0 | 0 | 2 | 8 | 0 | 0 | 6 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 805 | 95.03% |
| | E | 0 | 1 | 8 | 0 | 725 | 1 | 9 | 0 | 0 | 0 | 5 | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 4 | 0 | 5 | 768 | 94.40% |
| | F | 0 | 1 | 0 | 1 | 0 | 728 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 21 | 1 | 0 | 2 | 10 | 0 | 2 | 0 | 1 | 2 | 0 | 775 | 93.94% |
| | G | 0 | 4 | 5 | 5 | 11 | 0 | 736 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 773 | 95.21% |
| | H | 0 | 12 | 0 | 15 | 3 | 1 | 4 | 644 | 0 | 1 | 24 | 0 | 1 | 4 | 6 | 1 | 0 | 13 | 1 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 734 | 87.74% |
| | I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 729 | 25 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 755 | 96.56% |
| | J | 1 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 32 | 705 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 747 | 94.38% |
| | K | 0 | 3 | 0 | 0 | 10 | 0 | 2 | 25 | 0 | 0 | 671 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 739 | 90.80% |
| | L | 0 | 0 | 1 | 0 | 3 | 0 | 2 | 2 | 1 | 2 | 2 | 744 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 761 | 97.77% |
| | M | 1 | 7 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 767 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 6 | 3 | 0 | 0 | 0 | 792 | 96.84% |
| | N | 1 | 1 | 0 | 10 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 2 | 3 | 747 | 3 | 0 | 1 | 6 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 783 | 95.40% |
| | O | 0 | 0 | 2 | 10 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 725 | 0 | 9 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 753 | 96.28% |
| | P | 0 | 1 | 1 | 4 | 1 | 32 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 754 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 803 | 93.90% |
| | Q | 0 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 3 | 752 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 783 | 96.04% |
| | R | 0 | 19 | 0 | 2 | 0 | 2 | 0 | 10 | 0 | 0 | 15 | 2 | 0 | 6 | 0 | 0 | 1 | 701 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 758 | 92.48% |
| | S | 0 | 5 | 0 | 1 | 6 | 0 | 1 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 725 | 1 | 1 | 0 | 0 | 0 | 0 | 3 | 748 | 96.93% |
| | T | 0 | 3 | 2 | 1 | 0 | 3 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 769 | 0 | 0 | 0 | 0 | 12 | 1 | 796 | 96.61% |
| | U | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 7 | 0 | 0 | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 796 | 1 | 0 | 0 | 0 | 0 | 813 | 97.91% |
| | V | 0 | 12 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 740 | 1 | 0 | 2 | 0 | 764 | 96.86% |
| | W | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 3 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 740 | 0 | 0 | 0 | 752 | 98.40% |
| | X | 0 | 1 | 0 | 3 | 4 | 0 | 0 | 0 | 1 | 0 | 10 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 760 | 0 | 2 | 787 | 96.57% |
| | Y | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 8 | 0 | 3 | 1 | 1 | 765 | 0 | 786 | 97.33% |
| | Z | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 720 | 734 | 98.09% |
| Total | | 794 | 797 | 730 | 825 | 780 | 773 | 770 | 722 | 767 | 738 | 733 | 756 | 779 | 770 | 772 | 786 | 781 | 760 | 742 | 794 | 803 | 774 | 750 | 785 | 785 | 734 | 20000 | 95.59% |



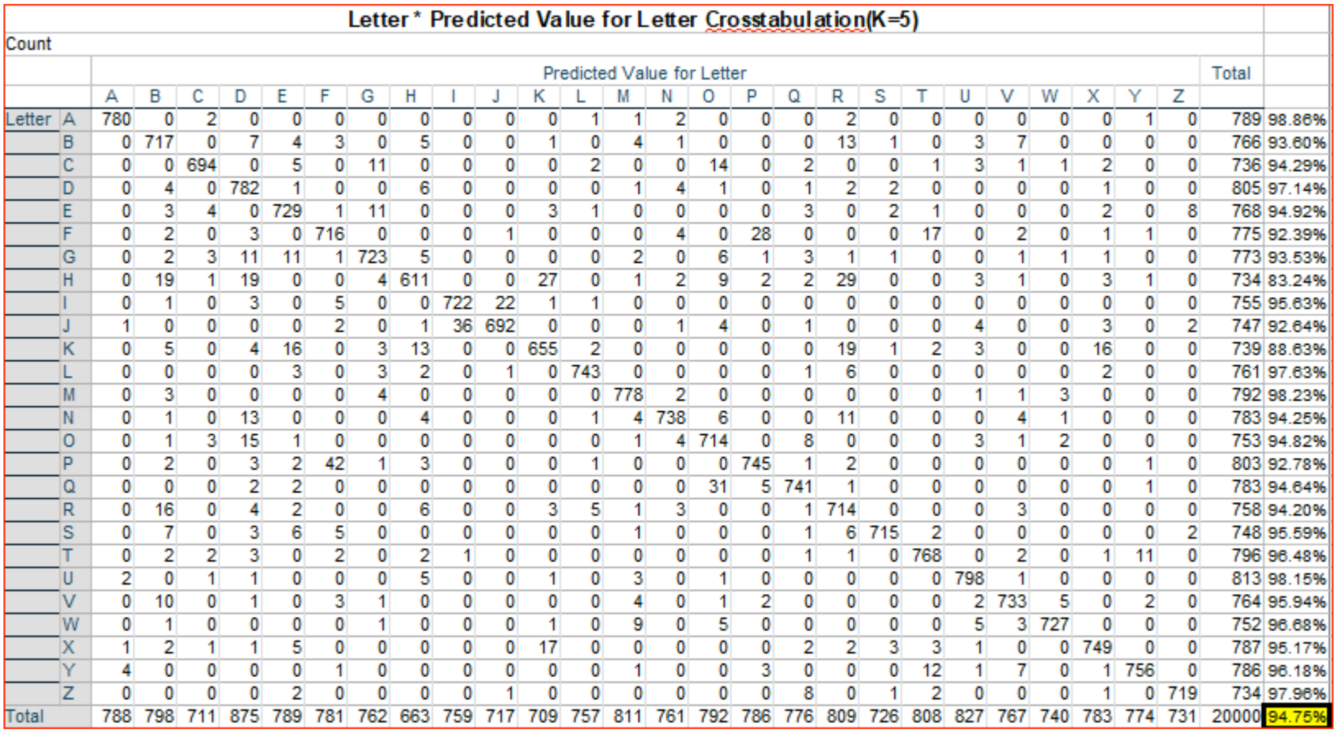
**Letter * Predicted Value for Letter Crosstabulation(K=3)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 779 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 789 | 98.73% |
| | B | 0 | 720 | 0 | 9 | 6 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 10 | 1 | 0 | 1 | 8 | 0 | 1 | 1 | 1 | 766 | 93.99% |
| | C | 1 | 0 | 695 | 0 | 7 | 0 | 11 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 8 | 0 | 2 | 0 | 0 | 1 | 7 | 0 | 3 | 0 | 0 | 0 | 736 | 94.43% |
| | D | 1 | 4 | 0 | 779 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 6 | 4 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 805 | 96.77% |
| | E | 0 | 3 | 4 | 0 | 720 | 4 | 12 | 0 | 0 | 0 | 2 | 5 | 0 | 0 | 0 | 1 | 2 | 0 | 2 | 1 | 1 | 0 | 0 | 4 | 0 | 7 | 768 | 93.75% |
| | F | 0 | 2 | 0 | 3 | 1 | 725 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 3 | 0 | 21 | 0 | 0 | 1 | 11 | 0 | 1 | 2 | 1 | 1 | 0 | 775 | 93.55% |
| | G | 0 | 7 | 3 | 10 | 9 | 0 | 725 | 3 | 0 | 0 | 0 | 0 | 5 | 0 | 2 | 1 | 2 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 773 | 93.79% |
| | H | 0 | 16 | 0 | 18 | 0 | 2 | 6 | 623 | 0 | 0 | 26 | 1 | 2 | 5 | 3 | 4 | 3 | 17 | 0 | 1 | 5 | 1 | 0 | 0 | 0 | 1 | 734 | 84.88% |
| | I | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 722 | 25 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 755 | 95.63% |
| | J | 0 | 1 | 0 | 2 | 0 | 3 | 0 | 2 | 31 | 693 | 0 | 1 | 0 | 2 | 1 | 0 | 2 | 0 | 1 | 0 | 4 | 0 | 0 | 3 | 0 | 1 | 747 | 92.77% |
| | K | 0 | 5 | 1 | 5 | 8 | 0 | 0 | 15 | 0 | 0 | 657 | 2 | 0 | 0 | 0 | 1 | 0 | 12 | 0 | 1 | 4 | 0 | 1 | 27 | 0 | 0 | 739 | 88.90% |
| | L | 0 | 0 | 0 | 0 | 4 | 0 | 7 | 0 | 0 | 1 | 2 | 742 | 0 | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 761 | 97.50% |
| | M | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 775 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 5 | 0 | 0 | 0 | 792 | 97.85% |
| | N | 1 | 2 | 0 | 11 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 2 | 4 | 743 | 4 | 0 | 1 | 5 | 0 | 2 | 0 | 4 | 1 | 0 | 0 | 0 | 783 | 94.89% |
| | O | 0 | 1 | 3 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 711 | 0 | 10 | 1 | 0 | 1 | 5 | 0 | 3 | 0 | 0 | 0 | 753 | 94.42% |
| | P | 0 | 1 | 0 | 0 | 0 | 40 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 750 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 803 | 93.40% |
| | Q | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 23 | 8 | 743 | 4 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 783 | 94.89% |
| | R | 0 | 24 | 0 | 6 | 1 | 1 | 1 | 3 | 0 | 0 | 6 | 4 | 1 | 17 | 0 | 1 | 2 | 688 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 758 | 90.77% |
| | S | 0 | 2 | 0 | 3 | 6 | 3 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 723 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 748 | 96.66% |
| | T | 0 | 1 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 775 | 0 | 0 | 0 | 0 | 13 | 0 | 796 | 97.36% |
| | U | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 804 | 0 | 0 | 0 | 0 | 0 | 813 | 98.89% |
| | V | 0 | 8 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 5 | 2 | 1 | 2 | 0 | 0 | 0 | 1 | 1 | 733 | 4 | 0 | 3 | 0 | 764 | 95.94% |
| | W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 4 | 0 | 0 | 1 | 0 | 0 | 7 | 3 | 732 | 0 | 1 | 0 | 752 | 97.34% |
| | X | 0 | 0 | 1 | 3 | 5 | 0 | 0 | 0 | 0 | 0 | 13 | 3 | 0 | 0 | 2 | 1 | 0 | 1 | 1 | 2 | 1 | 0 | 0 | 753 | 1 | 0 | 787 | 95.68% |
| | Y | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 11 | 2 | 2 | 0 | 1 | 765 | 0 | 786 | 97.33% |
| | Z | 0 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 8 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 716 | 734 | 97.55% |
| Total | | 787 | 798 | 711 | 865 | 772 | 786 | 767 | 665 | 755 | 722 | 709 | 768 | 804 | 789 | 764 | 791 | 780 | 748 | 734 | 813 | 849 | 758 | 752 | 794 | 791 | 728 | 20000 | 94.91% |

IS467
Homework#4
Date: 06-03-2017

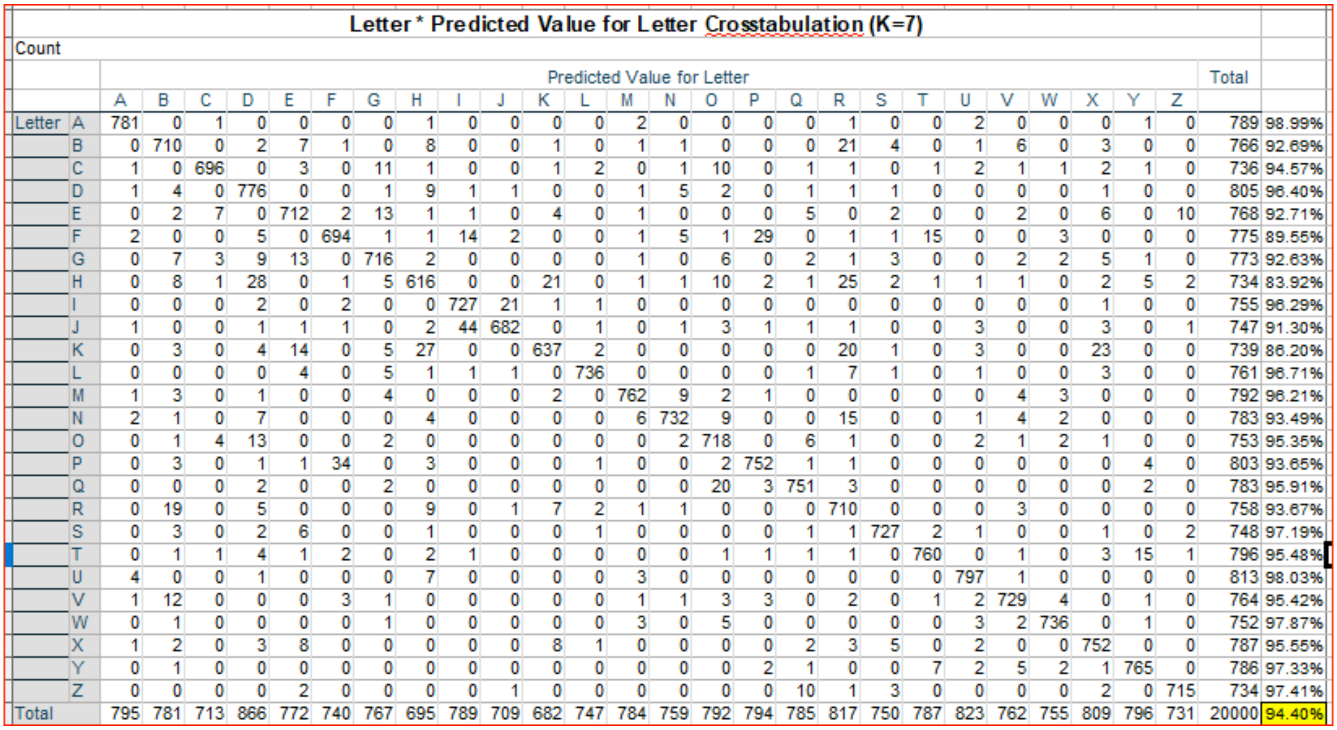
Name: Kai Chung, Ying
Email: kying@mail.depaul.edu



**Letter * Predicted Value for Letter Crosstabulation(K=5)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 780 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 789 | 98.86% |
| | B | 0 | 717 | 0 | 7 | 4 | 3 | 0 | 5 | 0 | 0 | 1 | 0 | 4 | 1 | 0 | 0 | 0 | 13 | 1 | 0 | 3 | 7 | 0 | 0 | 0 | 0 | 766 | 93.60% |
| | C | 0 | 0 | 694 | 0 | 5 | 0 | 11 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 14 | 0 | 2 | 0 | 0 | 1 | 3 | 1 | 1 | 2 | 0 | 0 | 736 | 94.29% |
| | D | 0 | 4 | 0 | 782 | 1 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 1 | 4 | 1 | 0 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 805 | 97.14% |
| | E | 0 | 3 | 4 | 0 | 729 | 1 | 11 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 2 | 1 | 0 | 0 | 0 | 2 | 0 | 8 | 768 | 94.92% |
| | F | 0 | 2 | 0 | 3 | 0 | 716 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 28 | 0 | 0 | 0 | 17 | 0 | 2 | 0 | 1 | 1 | 0 | 775 | 92.39% |
| | G | 0 | 2 | 3 | 11 | 11 | 1 | 723 | 5 | 0 | 0 | 0 | 0 | 2 | 0 | 6 | 1 | 3 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 773 | 93.53% |
| | H | 0 | 19 | 1 | 19 | 0 | 0 | 4 | 611 | 0 | 0 | 27 | 0 | 1 | 2 | 9 | 2 | 2 | 29 | 0 | 0 | 3 | 1 | 0 | 3 | 1 | 0 | 734 | 83.24% |
| | I | 0 | 1 | 0 | 3 | 0 | 5 | 0 | 0 | 722 | 22 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 755 | 95.63% |
| | J | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 36 | 692 | 0 | 0 | 0 | 1 | 4 | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 0 | 3 | 0 | 2 | 747 | 92.64% |
| | K | 0 | 5 | 0 | 4 | 16 | 0 | 3 | 13 | 0 | 0 | 655 | 2 | 0 | 0 | 0 | 0 | 0 | 19 | 1 | 2 | 3 | 0 | 0 | 16 | 0 | 0 | 739 | 88.63% |
| | L | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 2 | 0 | 1 | 0 | 743 | 0 | 0 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 761 | 97.63% |
| | M | 0 | 3 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 778 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 0 | 0 | 0 | 792 | 98.23% |
| | N | 0 | 1 | 0 | 13 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 1 | 4 | 738 | 6 | 0 | 0 | 11 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 0 | 783 | 94.25% |
| | O | 0 | 1 | 3 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 714 | 0 | 8 | 0 | 0 | 0 | 3 | 1 | 2 | 0 | 0 | 0 | 753 | 94.82% |
| | P | 0 | 2 | 0 | 3 | 2 | 42 | 1 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 745 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 803 | 92.78% |
| | Q | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 | 5 | 741 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 783 | 94.64% |
| | R | 0 | 16 | 0 | 4 | 2 | 0 | 0 | 6 | 0 | 0 | 3 | 5 | 1 | 3 | 0 | 0 | 1 | 714 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 758 | 94.20% |
| | S | 0 | 7 | 0 | 3 | 6 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 6 | 715 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 748 | 95.59% |
| | T | 0 | 2 | 2 | 3 | 0 | 2 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 768 | 0 | 2 | 0 | 1 | 11 | 0 | 796 | 96.48% |
| | U | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 5 | 0 | 0 | 1 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 798 | 1 | 0 | 0 | 0 | 0 | 813 | 98.15% |
| | V | 0 | 10 | 0 | 1 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 2 | 733 | 5 | 0 | 2 | 0 | 764 | 95.94% |
| | W | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 9 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 5 | 3 | 727 | 0 | 0 | 0 | 752 | 96.68% |
| | X | 1 | 2 | 1 | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 3 | 3 | 1 | 0 | 0 | 749 | 0 | 0 | 787 | 95.17% |
| | Y | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 0 | 0 | 0 | 12 | 1 | 7 | 0 | 1 | 756 | 0 | 786 | 96.18% |
| | Z | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 719 | 734 | 97.96% |
| Total | | 788 | 798 | 711 | 875 | 789 | 781 | 762 | 663 | 759 | 717 | 709 | 757 | 811 | 761 | 792 | 786 | 776 | 809 | 726 | 808 | 827 | 767 | 740 | 783 | 774 | 731 | 20000 | 94.75% |



**Letter * Predicted Value for Letter Crosstabulation (K=7)**

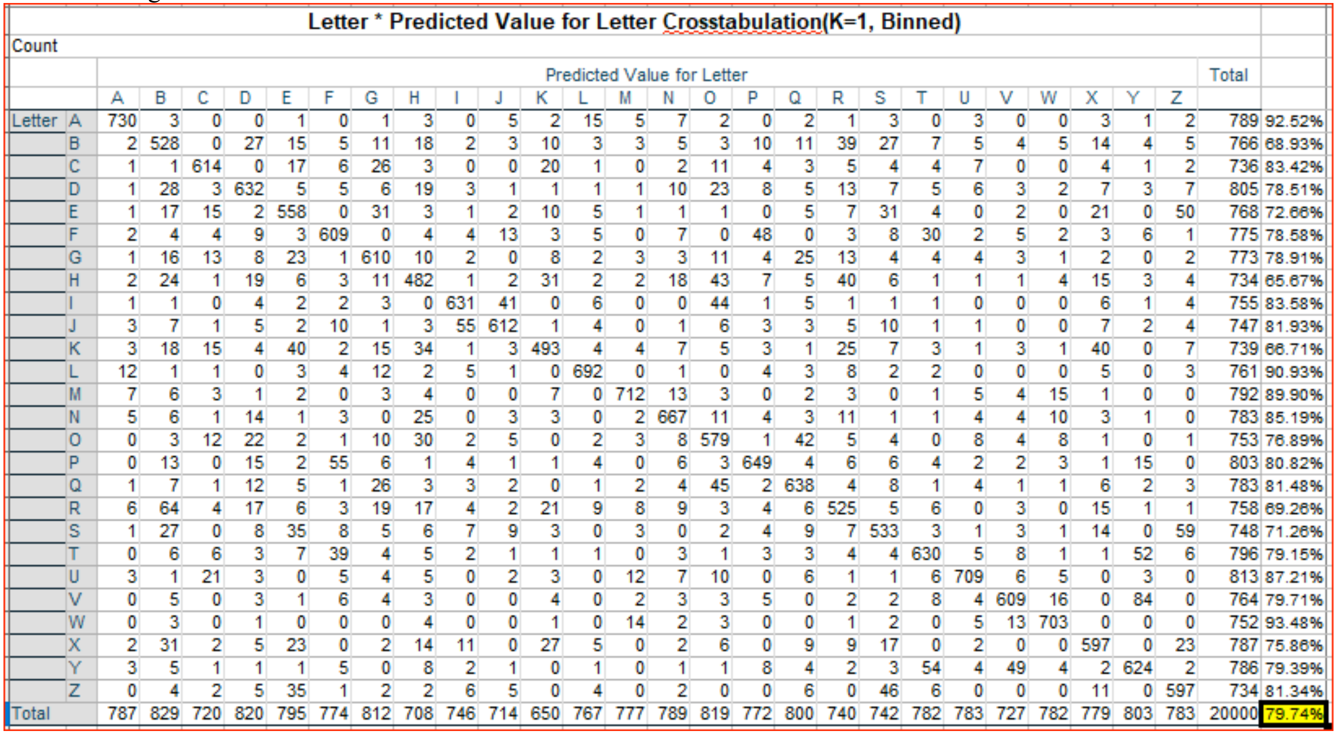
Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 781 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 789 | 98.99% |
| | B | 0 | 710 | 0 | 2 | 7 | 1 | 0 | 8 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 21 | 4 | 0 | 1 | 6 | 0 | 3 | 0 | 0 | 766 | 92.69% |
| | C | 1 | 0 | 696 | 0 | 3 | 0 | 11 | 1 | 0 | 0 | 1 | 2 | 0 | 1 | 10 | 0 | 1 | 1 | 0 | 1 | 2 | 1 | 1 | 2 | 1 | 0 | 736 | 94.57% |
| | D | 1 | 4 | 0 | 776 | 0 | 0 | 1 | 9 | 1 | 1 | 0 | 0 | 1 | 5 | 2 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 805 | 96.40% |
| | E | 0 | 2 | 7 | 0 | 712 | 2 | 13 | 1 | 1 | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 2 | 0 | 0 | 2 | 0 | 6 | 0 | 10 | 768 | 92.71% |
| | F | 2 | 0 | 0 | 5 | 0 | 694 | 1 | 1 | 14 | 2 | 0 | 0 | 1 | 5 | 1 | 29 | 0 | 1 | 1 | 15 | 0 | 0 | 3 | 0 | 0 | 0 | 775 | 89.55% |
| | G | 0 | 7 | 3 | 9 | 13 | 0 | 716 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 6 | 0 | 2 | 1 | 3 | 0 | 0 | 2 | 2 | 5 | 1 | 0 | 773 | 92.63% |
| | H | 0 | 8 | 1 | 28 | 0 | 1 | 5 | 616 | 0 | 0 | 21 | 0 | 1 | 1 | 10 | 2 | 1 | 25 | 2 | 1 | 1 | 1 | 0 | 2 | 5 | 2 | 734 | 83.92% |
| | I | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 727 | 21 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 755 | 96.29% |
| | J | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 2 | 44 | 682 | 0 | 1 | 0 | 1 | 3 | 1 | 1 | 1 | 0 | 0 | 3 | 0 | 0 | 3 | 0 | 1 | 747 | 91.30% |
| | K | 0 | 3 | 0 | 4 | 14 | 0 | 5 | 27 | 0 | 0 | 637 | 2 | 0 | 0 | 0 | 0 | 0 | 20 | 1 | 0 | 3 | 0 | 0 | 23 | 0 | 0 | 739 | 86.20% |
| | L | 0 | 0 | 0 | 0 | 4 | 0 | 5 | 1 | 1 | 1 | 0 | 736 | 0 | 0 | 0 | 0 | 1 | 7 | 1 | 0 | 1 | 0 | 0 | 3 | 0 | 0 | 761 | 96.71% |
| | M | 1 | 3 | 0 | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 2 | 0 | 762 | 9 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 3 | 0 | 0 | 0 | 792 | 96.21% |
| | N | 2 | 1 | 0 | 7 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 6 | 732 | 9 | 0 | 0 | 15 | 0 | 0 | 1 | 4 | 2 | 0 | 0 | 0 | 783 | 93.49% |
| | O | 0 | 1 | 4 | 13 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 718 | 0 | 6 | 1 | 0 | 0 | 2 | 1 | 2 | 1 | 0 | 0 | 753 | 95.35% |
| | P | 0 | 3 | 0 | 1 | 1 | 34 | 0 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 752 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 803 | 93.65% |
| | Q | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 3 | 751 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 783 | 95.91% |
| | R | 0 | 19 | 0 | 5 | 0 | 0 | 0 | 9 | 0 | 1 | 7 | 2 | 1 | 1 | 0 | 0 | 0 | 710 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 758 | 93.67% |
| | S | 0 | 3 | 0 | 2 | 6 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 727 | 2 | 1 | 0 | 0 | 1 | 0 | 2 | 748 | 97.19% |
| | T | 0 | 1 | 1 | 4 | 1 | 2 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 760 | 0 | 1 | 0 | 3 | 15 | 1 | 796 | 95.48% |
| | U | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 797 | 1 | 0 | 0 | 0 | 0 | 813 | 98.03% |
| | V | 1 | 12 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 3 | 0 | 2 | 0 | 1 | 2 | 729 | 4 | 0 | 1 | 0 | 764 | 95.42% |
| | W | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 3 | 2 | 736 | 0 | 1 | 0 | 752 | 97.87% |
| | X | 1 | 2 | 0 | 3 | 8 | 0 | 0 | 0 | 0 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 0 | 2 | 0 | 0 | 752 | 0 | 0 | 787 | 95.55% |
| | Y | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 7 | 2 | 5 | 2 | 1 | 765 | 0 | 786 | 97.33% |
| | Z | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 1 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 715 | 734 | 97.41% |
| Total | | 795 | 781 | 713 | 866 | 772 | 740 | 767 | 695 | 789 | 709 | 682 | 747 | 784 | 759 | 792 | 794 | 785 | 817 | 750 | 787 | 823 | 762 | 755 | 809 | 796 | 731 | 20000 | 94.40% |

IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

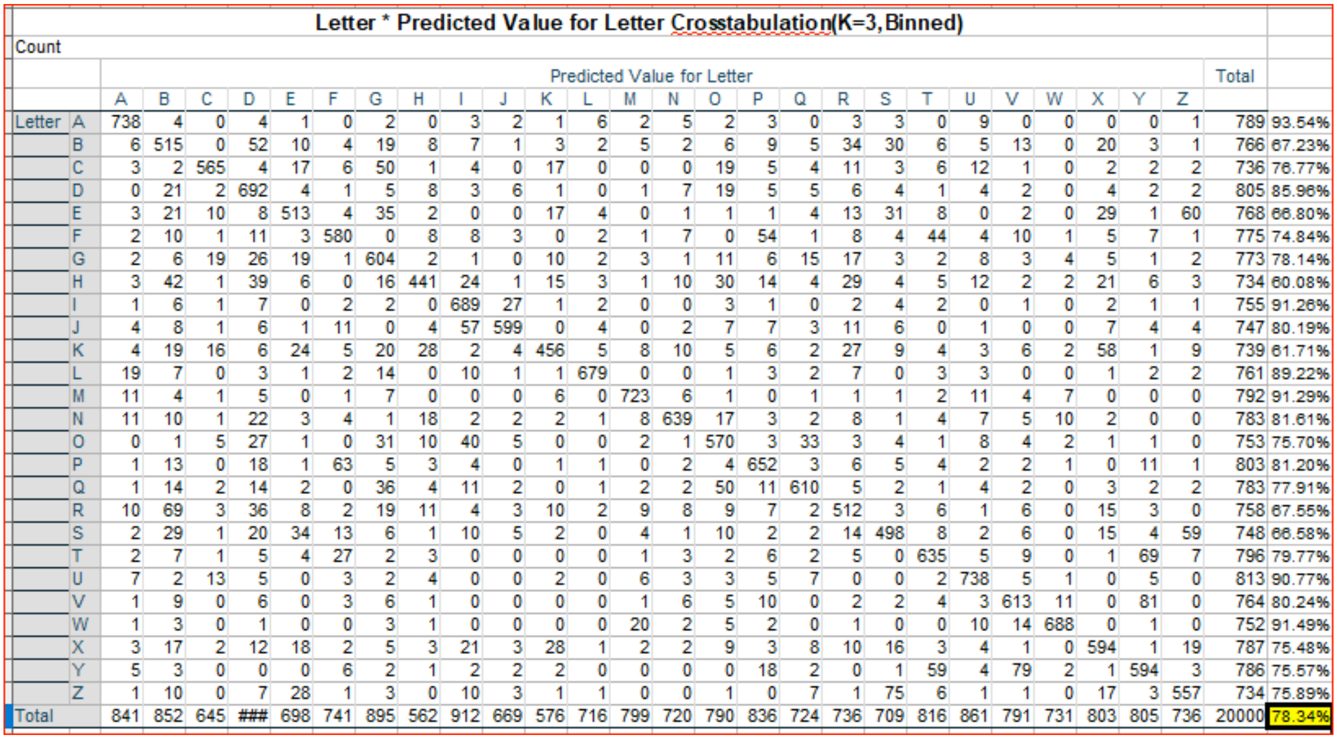
The Following is the binned data result:



**Letter * Predicted Value for Letter Crosstabulation(K=1, Binned)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 730 | 3 | 0 | 0 | 1 | 0 | 1 | 3 | 0 | 5 | 2 | 15 | 5 | 7 | 2 | 0 | 2 | 1 | 3 | 0 | 3 | 0 | 0 | 3 | 1 | 2 | 789 | 92.52% |
| | B | 2 | 528 | 0 | 27 | 15 | 5 | 11 | 18 | 2 | 3 | 10 | 3 | 3 | 5 | 3 | 10 | 11 | 39 | 27 | 7 | 5 | 4 | 5 | 14 | 4 | 5 | 766 | 68.93% |
| | C | 1 | 1 | 614 | 0 | 17 | 6 | 26 | 3 | 0 | 0 | 20 | 1 | 0 | 2 | 11 | 4 | 3 | 5 | 4 | 4 | 7 | 0 | 0 | 4 | 1 | 2 | 736 | 83.42% |
| | D | 1 | 28 | 3 | 632 | 5 | 5 | 6 | 19 | 3 | 1 | 1 | 1 | 1 | 10 | 23 | 8 | 5 | 13 | 7 | 5 | 6 | 3 | 2 | 7 | 3 | 7 | 805 | 78.51% |
| | E | 1 | 17 | 15 | 2 | 558 | 0 | 31 | 3 | 1 | 2 | 10 | 5 | 1 | 1 | 1 | 0 | 5 | 7 | 31 | 4 | 0 | 2 | 0 | 21 | 0 | 50 | 768 | 72.66% |
| | F | 2 | 4 | 4 | 9 | 3 | 609 | 0 | 4 | 4 | 13 | 3 | 5 | 0 | 7 | 0 | 48 | 0 | 3 | 8 | 30 | 2 | 5 | 2 | 3 | 6 | 1 | 775 | 78.58% |
| | G | 1 | 16 | 13 | 8 | 23 | 1 | 610 | 10 | 2 | 0 | 8 | 2 | 3 | 3 | 11 | 4 | 25 | 13 | 4 | 4 | 4 | 3 | 1 | 2 | 0 | 2 | 773 | 78.91% |
| | H | 2 | 24 | 1 | 19 | 6 | 3 | 11 | 482 | 1 | 2 | 31 | 2 | 2 | 18 | 43 | 7 | 5 | 40 | 6 | 1 | 1 | 1 | 4 | 15 | 3 | 4 | 734 | 65.67% |
| | I | 1 | 1 | 0 | 4 | 2 | 2 | 3 | 0 | 631 | 41 | 0 | 6 | 0 | 0 | 44 | 1 | 5 | 1 | 1 | 1 | 0 | 0 | 0 | 6 | 1 | 4 | 755 | 83.58% |
| | J | 3 | 7 | 1 | 5 | 2 | 10 | 1 | 3 | 55 | 612 | 1 | 4 | 0 | 1 | 6 | 3 | 3 | 5 | 10 | 1 | 1 | 0 | 0 | 7 | 2 | 4 | 747 | 81.93% |
| | K | 3 | 18 | 15 | 4 | 40 | 2 | 15 | 34 | 1 | 3 | 493 | 4 | 4 | 7 | 5 | 3 | 1 | 25 | 7 | 3 | 1 | 3 | 1 | 40 | 0 | 7 | 739 | 66.71% |
| | L | 12 | 1 | 1 | 0 | 3 | 4 | 12 | 2 | 5 | 1 | 0 | 692 | 0 | 1 | 0 | 4 | 3 | 8 | 2 | 2 | 0 | 0 | 0 | 5 | 0 | 3 | 761 | 90.93% |
| | M | 7 | 6 | 3 | 1 | 2 | 0 | 3 | 4 | 0 | 0 | 7 | 0 | 712 | 13 | 3 | 0 | 2 | 3 | 0 | 1 | 5 | 4 | 15 | 1 | 0 | 0 | 792 | 89.90% |
| | N | 5 | 6 | 1 | 14 | 1 | 3 | 0 | 25 | 0 | 3 | 3 | 0 | 2 | 667 | 11 | 4 | 3 | 11 | 1 | 1 | 4 | 4 | 10 | 3 | 1 | 0 | 783 | 85.19% |
| | O | 0 | 3 | 12 | 22 | 2 | 1 | 10 | 30 | 2 | 5 | 0 | 2 | 3 | 8 | 579 | 1 | 42 | 5 | 4 | 0 | 8 | 4 | 8 | 1 | 0 | 1 | 753 | 76.89% |
| | P | 0 | 13 | 0 | 15 | 2 | 55 | 6 | 1 | 4 | 1 | 1 | 4 | 0 | 6 | 3 | 649 | 4 | 6 | 6 | 4 | 2 | 2 | 3 | 1 | 15 | 0 | 803 | 80.82% |
| | Q | 1 | 7 | 1 | 12 | 5 | 1 | 26 | 3 | 3 | 2 | 0 | 1 | 2 | 4 | 45 | 2 | 638 | 4 | 8 | 1 | 4 | 1 | 1 | 6 | 2 | 3 | 783 | 81.48% |
| | R | 6 | 64 | 4 | 17 | 6 | 3 | 19 | 17 | 4 | 2 | 21 | 9 | 8 | 9 | 3 | 4 | 6 | 525 | 5 | 6 | 0 | 3 | 0 | 15 | 1 | 1 | 758 | 69.26% |
| | S | 1 | 27 | 0 | 8 | 35 | 8 | 5 | 6 | 7 | 9 | 3 | 0 | 3 | 0 | 2 | 4 | 9 | 7 | 533 | 3 | 1 | 3 | 1 | 14 | 0 | 59 | 748 | 71.26% |
| | T | 0 | 6 | 6 | 3 | 7 | 39 | 4 | 5 | 2 | 1 | 1 | 1 | 0 | 3 | 1 | 3 | 3 | 4 | 4 | 630 | 5 | 8 | 1 | 1 | 52 | 6 | 796 | 79.15% |
| | U | 3 | 1 | 21 | 3 | 0 | 5 | 4 | 5 | 0 | 2 | 3 | 0 | 12 | 7 | 10 | 0 | 6 | 1 | 1 | 6 | 709 | 6 | 5 | 0 | 3 | 0 | 813 | 87.21% |
| | V | 0 | 5 | 0 | 3 | 1 | 6 | 4 | 3 | 0 | 0 | 4 | 0 | 2 | 3 | 3 | 5 | 0 | 2 | 2 | 8 | 4 | 609 | 16 | 0 | 84 | 0 | 764 | 79.71% |
| | W | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 0 | 1 | 0 | 14 | 2 | 3 | 0 | 0 | 1 | 2 | 0 | 5 | 13 | 703 | 0 | 0 | 0 | 752 | 93.48% |
| | X | 2 | 31 | 2 | 5 | 23 | 0 | 2 | 14 | 11 | 0 | 27 | 5 | 0 | 2 | 6 | 0 | 9 | 9 | 17 | 0 | 2 | 0 | 0 | 597 | 0 | 23 | 787 | 75.86% |
| | Y | 3 | 5 | 1 | 1 | 1 | 5 | 0 | 8 | 2 | 1 | 0 | 1 | 0 | 1 | 1 | 8 | 4 | 2 | 3 | 54 | 4 | 49 | 4 | 2 | 624 | 2 | 786 | 79.39% |
| | Z | 0 | 4 | 2 | 5 | 35 | 1 | 2 | 2 | 6 | 5 | 0 | 4 | 0 | 2 | 0 | 0 | 6 | 0 | 46 | 6 | 0 | 0 | 0 | 11 | 0 | 597 | 734 | 81.34% |
| Total | | 787 | 829 | 720 | 820 | 795 | 774 | 812 | 708 | 746 | 714 | 650 | 767 | 777 | 789 | 819 | 772 | 800 | 740 | 742 | 782 | 783 | 727 | 782 | 779 | 803 | 783 | 20000 | 79.74% |



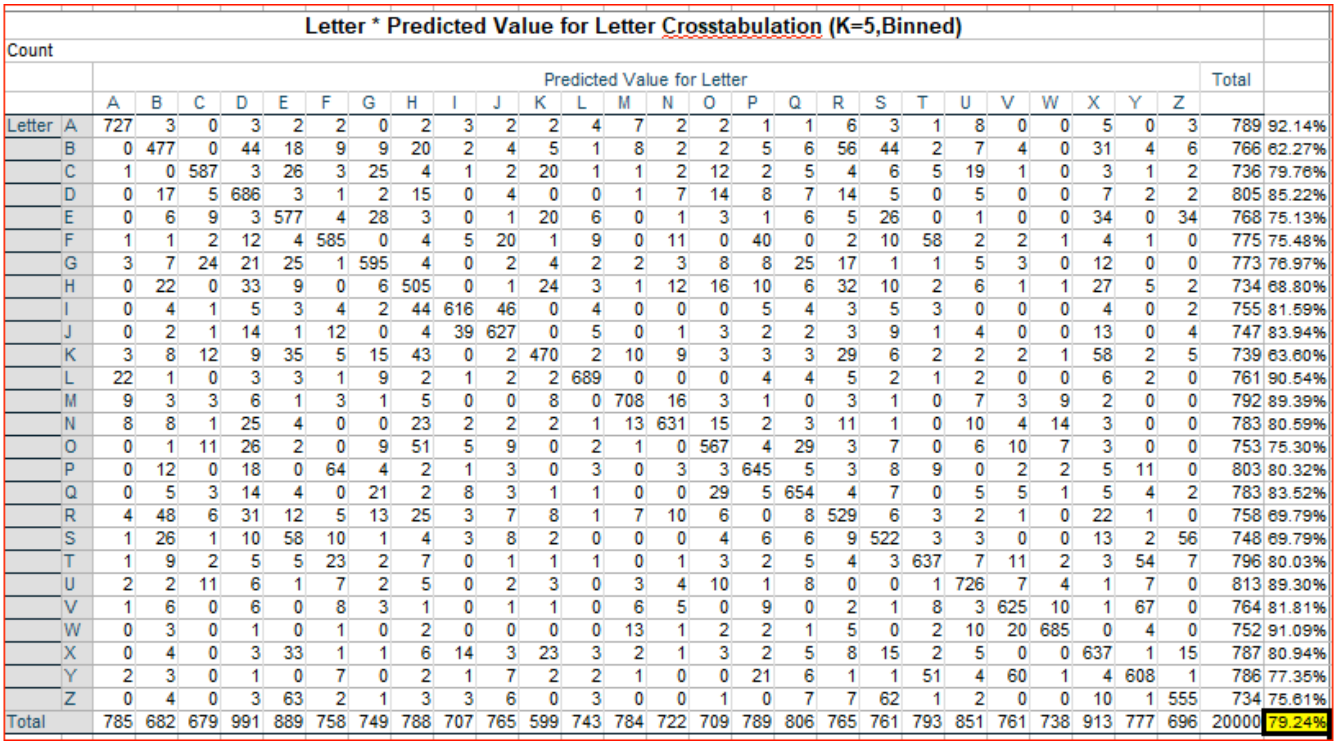
**Letter * Predicted Value for Letter Crosstabulation(K=3,Binned)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 738 | 4 | 0 | 4 | 1 | 0 | 2 | 0 | 3 | 2 | 1 | 6 | 2 | 5 | 2 | 3 | 0 | 3 | 3 | 0 | 9 | 0 | 0 | 0 | 0 | 1 | 789 | 93.54% |
| | B | 6 | 515 | 0 | 52 | 10 | 4 | 19 | 8 | 7 | 1 | 3 | 2 | 5 | 2 | 6 | 9 | 5 | 34 | 30 | 6 | 5 | 13 | 0 | 20 | 3 | 1 | 766 | 67.23% |
| | C | 3 | 2 | 565 | 4 | 17 | 6 | 50 | 1 | 4 | 0 | 17 | 0 | 0 | 0 | 19 | 5 | 4 | 11 | 3 | 6 | 12 | 1 | 0 | 2 | 2 | 2 | 736 | 76.77% |
| | D | 0 | 21 | 2 | 692 | 4 | 1 | 5 | 8 | 3 | 6 | 1 | 0 | 1 | 7 | 19 | 5 | 5 | 6 | 4 | 1 | 4 | 2 | 0 | 4 | 2 | 2 | 805 | 85.96% |
| | E | 3 | 21 | 10 | 8 | 513 | 4 | 35 | 2 | 0 | 0 | 17 | 4 | 0 | 1 | 1 | 1 | 4 | 13 | 31 | 8 | 0 | 2 | 0 | 29 | 1 | 60 | 768 | 66.80% |
| | F | 2 | 10 | 1 | 11 | 3 | 580 | 0 | 8 | 8 | 3 | 0 | 2 | 1 | 7 | 0 | 54 | 1 | 8 | 4 | 44 | 4 | 10 | 1 | 5 | 7 | 1 | 775 | 74.84% |
| | G | 2 | 6 | 19 | 26 | 19 | 1 | 604 | 2 | 1 | 0 | 10 | 2 | 3 | 1 | 11 | 6 | 15 | 17 | 3 | 2 | 8 | 3 | 4 | 5 | 1 | 2 | 773 | 78.14% |
| | H | 3 | 42 | 1 | 39 | 6 | 0 | 16 | 441 | 24 | 1 | 15 | 3 | 1 | 10 | 30 | 14 | 4 | 29 | 4 | 5 | 12 | 2 | 2 | 21 | 6 | 3 | 734 | 60.08% |
| | I | 1 | 6 | 1 | 7 | 0 | 2 | 2 | 0 | 689 | 27 | 1 | 2 | 0 | 0 | 3 | 1 | 0 | 2 | 4 | 2 | 0 | 1 | 0 | 2 | 1 | 1 | 755 | 91.26% |
| | J | 4 | 8 | 1 | 6 | 1 | 11 | 0 | 4 | 57 | 599 | 0 | 4 | 0 | 2 | 7 | 7 | 3 | 11 | 6 | 0 | 1 | 0 | 0 | 7 | 4 | 4 | 747 | 80.19% |
| | K | 4 | 19 | 16 | 6 | 24 | 5 | 20 | 28 | 2 | 4 | 456 | 5 | 8 | 10 | 5 | 6 | 2 | 27 | 9 | 4 | 3 | 6 | 2 | 58 | 1 | 9 | 739 | 61.71% |
| | L | 19 | 7 | 0 | 3 | 1 | 2 | 14 | 0 | 10 | 1 | 1 | 679 | 0 | 0 | 1 | 3 | 2 | 7 | 0 | 3 | 3 | 0 | 0 | 1 | 2 | 2 | 761 | 89.22% |
| | M | 11 | 4 | 1 | 5 | 0 | 1 | 7 | 0 | 0 | 0 | 6 | 0 | 723 | 6 | 1 | 0 | 1 | 1 | 1 | 2 | 11 | 4 | 7 | 0 | 0 | 0 | 792 | 91.29% |
| | N | 11 | 10 | 1 | 22 | 3 | 4 | 1 | 18 | 2 | 2 | 2 | 1 | 8 | 639 | 17 | 3 | 2 | 8 | 1 | 4 | 7 | 5 | 10 | 2 | 0 | 0 | 783 | 81.61% |
| | O | 0 | 1 | 5 | 27 | 1 | 0 | 31 | 10 | 40 | 5 | 0 | 0 | 2 | 1 | 570 | 3 | 33 | 3 | 4 | 1 | 8 | 4 | 2 | 1 | 1 | 0 | 753 | 75.70% |
| | P | 1 | 13 | 0 | 18 | 1 | 63 | 5 | 3 | 4 | 0 | 1 | 1 | 0 | 2 | 4 | 652 | 3 | 6 | 5 | 4 | 2 | 2 | 1 | 0 | 11 | 1 | 803 | 81.20% |
| | Q | 1 | 14 | 2 | 14 | 2 | 0 | 36 | 4 | 11 | 2 | 0 | 1 | 2 | 2 | 50 | 11 | 610 | 5 | 2 | 1 | 4 | 2 | 0 | 3 | 2 | 2 | 783 | 77.91% |
| | R | 10 | 69 | 3 | 36 | 8 | 2 | 19 | 11 | 4 | 3 | 10 | 2 | 9 | 8 | 9 | 7 | 2 | 512 | 3 | 6 | 1 | 6 | 0 | 15 | 3 | 0 | 758 | 67.55% |
| | S | 2 | 29 | 1 | 20 | 34 | 13 | 6 | 1 | 10 | 5 | 2 | 0 | 4 | 1 | 10 | 2 | 2 | 14 | 498 | 8 | 2 | 6 | 0 | 15 | 4 | 59 | 748 | 66.58% |
| | T | 2 | 7 | 1 | 5 | 4 | 27 | 2 | 3 | 0 | 0 | 0 | 0 | 1 | 3 | 2 | 6 | 2 | 5 | 0 | 635 | 5 | 9 | 0 | 1 | 69 | 7 | 796 | 79.77% |
| | U | 7 | 2 | 13 | 5 | 0 | 3 | 2 | 4 | 0 | 0 | 2 | 0 | 6 | 3 | 3 | 5 | 7 | 0 | 0 | 2 | 738 | 5 | 1 | 0 | 5 | 0 | 813 | 90.77% |
| | V | 1 | 9 | 0 | 6 | 0 | 3 | 6 | 1 | 0 | 0 | 0 | 0 | 1 | 6 | 5 | 10 | 0 | 2 | 2 | 4 | 3 | 613 | 11 | 0 | 81 | 0 | 764 | 80.24% |
| | W | 1 | 3 | 0 | 1 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 20 | 2 | 5 | 2 | 0 | 1 | 0 | 0 | 10 | 14 | 688 | 0 | 1 | 0 | 752 | 91.49% |
| | X | 3 | 17 | 2 | 12 | 18 | 2 | 5 | 3 | 21 | 3 | 28 | 1 | 2 | 2 | 9 | 3 | 8 | 10 | 16 | 3 | 4 | 1 | 0 | 594 | 1 | 19 | 787 | 75.48% |
| | Y | 5 | 3 | 0 | 0 | 0 | 6 | 2 | 1 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 18 | 2 | 0 | 1 | 59 | 4 | 79 | 2 | 1 | 594 | 3 | 786 | 75.57% |
| | Z | 1 | 10 | 0 | 7 | 28 | 1 | 3 | 0 | 10 | 3 | 1 | 1 | 0 | 0 | 1 | 0 | 7 | 1 | 75 | 6 | 1 | 1 | 0 | 17 | 3 | 557 | 734 | 75.89% |
| Total | | 841 | 852 | 645 | ### | 698 | 741 | 895 | 562 | 912 | 669 | 576 | 716 | 799 | 720 | 790 | 836 | 724 | 736 | 709 | 816 | 861 | 791 | 731 | 803 | 805 | 736 | 20000 | 78.34% |

IS467
Homework#4
Date: 06-03-2017

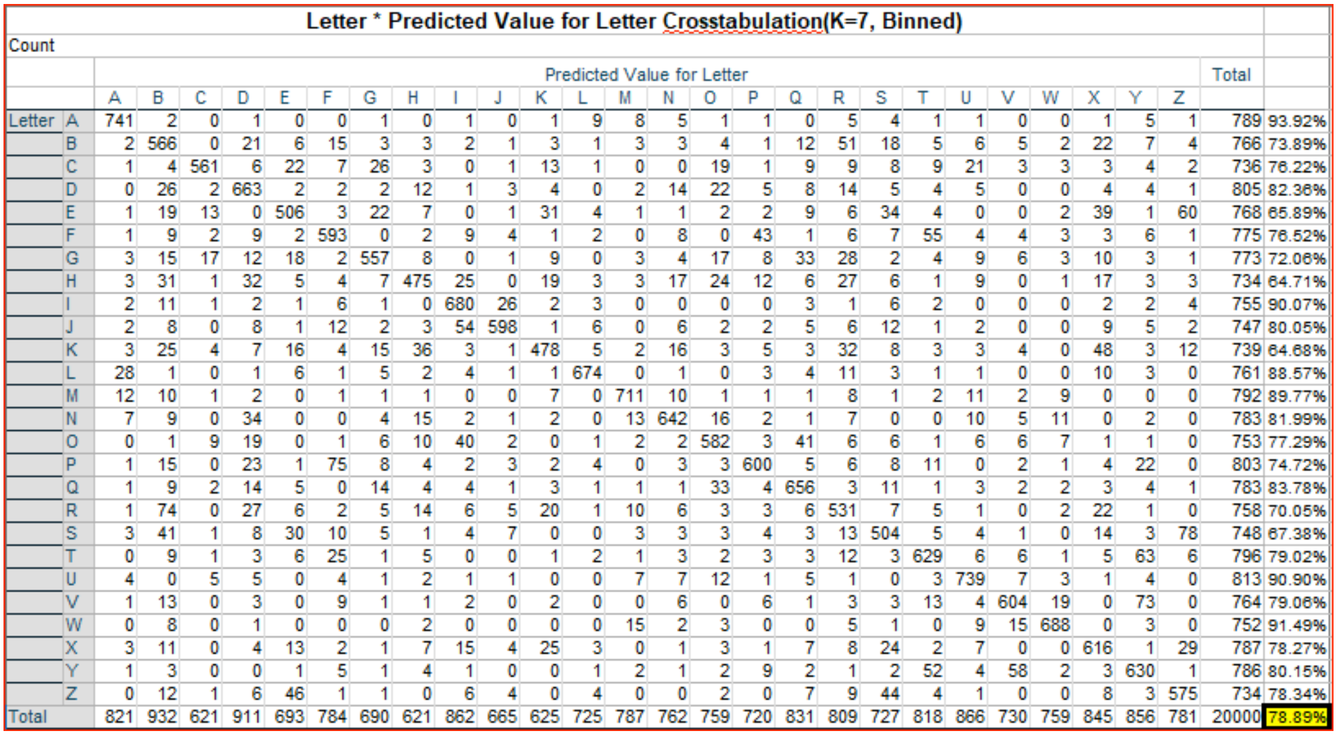
Name: Kai Chung, Ying
Email: kying@mail.depaul.edu



**Letter * Predicted Value for Letter Crosstabulation (K=5,Binned)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 727 | 3 | 0 | 3 | 2 | 2 | 0 | 2 | 3 | 2 | 2 | 4 | 7 | 2 | 2 | 1 | 1 | 6 | 3 | 1 | 8 | 0 | 0 | 5 | 0 | 3 | 789 | 92.14% |
| | B | 0 | 477 | 0 | 44 | 18 | 9 | 9 | 20 | 2 | 4 | 5 | 1 | 8 | 2 | 2 | 5 | 6 | 56 | 44 | 2 | 7 | 4 | 0 | 31 | 4 | 6 | 766 | 62.27% |
| | C | 1 | 0 | 587 | 3 | 26 | 3 | 25 | 4 | 1 | 2 | 20 | 1 | 1 | 2 | 12 | 2 | 5 | 4 | 6 | 5 | 19 | 1 | 0 | 3 | 1 | 2 | 736 | 79.76% |
| | D | 0 | 17 | 5 | 686 | 3 | 1 | 2 | 15 | 0 | 4 | 0 | 0 | 1 | 7 | 14 | 8 | 7 | 14 | 5 | 0 | 5 | 0 | 0 | 7 | 2 | 2 | 805 | 85.22% |
| | E | 0 | 6 | 9 | 3 | 577 | 4 | 28 | 3 | 0 | 1 | 20 | 6 | 0 | 1 | 3 | 1 | 6 | 5 | 26 | 0 | 1 | 0 | 0 | 34 | 0 | 34 | 768 | 75.13% |
| | F | 1 | 1 | 2 | 12 | 4 | 585 | 0 | 4 | 5 | 20 | 1 | 9 | 0 | 11 | 0 | 40 | 0 | 2 | 10 | 58 | 2 | 2 | 1 | 4 | 1 | 0 | 775 | 75.48% |
| | G | 3 | 7 | 24 | 21 | 25 | 1 | 595 | 4 | 0 | 2 | 4 | 2 | 2 | 3 | 8 | 8 | 25 | 17 | 1 | 1 | 5 | 3 | 0 | 12 | 0 | 0 | 773 | 76.97% |
| | H | 0 | 22 | 0 | 33 | 9 | 0 | 6 | 505 | 0 | 1 | 24 | 3 | 1 | 12 | 16 | 10 | 6 | 32 | 10 | 2 | 6 | 1 | 1 | 27 | 5 | 2 | 734 | 68.80% |
| | I | 0 | 4 | 1 | 5 | 3 | 4 | 2 | 44 | 616 | 46 | 0 | 4 | 0 | 0 | 0 | 5 | 4 | 3 | 5 | 3 | 0 | 0 | 0 | 4 | 0 | 2 | 755 | 81.59% |
| | J | 0 | 2 | 1 | 14 | 1 | 12 | 0 | 4 | 39 | 627 | 0 | 5 | 0 | 1 | 3 | 2 | 2 | 3 | 9 | 1 | 4 | 0 | 0 | 13 | 0 | 4 | 747 | 83.94% |
| | K | 3 | 8 | 12 | 9 | 35 | 5 | 15 | 43 | 0 | 2 | 470 | 2 | 10 | 9 | 3 | 3 | 3 | 29 | 6 | 2 | 2 | 2 | 1 | 58 | 2 | 5 | 739 | 63.60% |
| | L | 22 | 1 | 0 | 3 | 3 | 1 | 9 | 2 | 1 | 2 | 2 | 689 | 0 | 0 | 0 | 4 | 4 | 5 | 2 | 1 | 2 | 0 | 0 | 6 | 2 | 0 | 761 | 90.54% |
| | M | 9 | 3 | 3 | 6 | 1 | 3 | 1 | 5 | 0 | 0 | 8 | 0 | 708 | 16 | 3 | 1 | 0 | 3 | 1 | 0 | 7 | 3 | 9 | 2 | 0 | 0 | 792 | 89.39% |
| | N | 8 | 8 | 1 | 25 | 4 | 0 | 0 | 23 | 2 | 2 | 2 | 1 | 13 | 631 | 15 | 2 | 3 | 11 | 1 | 0 | 10 | 4 | 14 | 3 | 0 | 0 | 783 | 80.59% |
| | O | 0 | 1 | 11 | 26 | 2 | 0 | 9 | 51 | 5 | 9 | 0 | 2 | 1 | 0 | 567 | 4 | 29 | 3 | 7 | 0 | 6 | 10 | 7 | 3 | 0 | 0 | 753 | 75.30% |
| | P | 0 | 12 | 0 | 18 | 0 | 64 | 4 | 2 | 1 | 3 | 0 | 3 | 0 | 3 | 3 | 645 | 5 | 3 | 8 | 9 | 0 | 2 | 2 | 5 | 11 | 0 | 803 | 80.32% |
| | Q | 0 | 5 | 3 | 14 | 4 | 0 | 21 | 2 | 8 | 3 | 1 | 1 | 0 | 0 | 29 | 5 | 654 | 4 | 7 | 0 | 5 | 5 | 1 | 5 | 4 | 2 | 783 | 83.52% |
| | R | 4 | 48 | 6 | 31 | 12 | 5 | 13 | 25 | 3 | 7 | 8 | 1 | 7 | 10 | 6 | 0 | 8 | 529 | 6 | 3 | 2 | 1 | 0 | 22 | 1 | 0 | 758 | 69.79% |
| | S | 1 | 26 | 1 | 10 | 58 | 10 | 1 | 4 | 3 | 8 | 2 | 0 | 0 | 0 | 4 | 6 | 6 | 9 | 522 | 3 | 3 | 0 | 0 | 13 | 2 | 56 | 748 | 69.79% |
| | T | 1 | 9 | 2 | 5 | 5 | 23 | 2 | 7 | 0 | 1 | 1 | 1 | 0 | 1 | 3 | 2 | 5 | 4 | 3 | 637 | 7 | 11 | 2 | 3 | 54 | 7 | 796 | 80.03% |
| | U | 2 | 2 | 11 | 6 | 1 | 7 | 2 | 5 | 0 | 2 | 3 | 0 | 3 | 4 | 10 | 1 | 8 | 0 | 0 | 1 | 726 | 7 | 4 | 1 | 7 | 0 | 813 | 89.30% |
| | V | 1 | 6 | 0 | 6 | 0 | 8 | 3 | 1 | 0 | 1 | 1 | 0 | 6 | 5 | 0 | 9 | 0 | 2 | 1 | 8 | 3 | 625 | 10 | 1 | 67 | 0 | 764 | 81.81% |
| | W | 0 | 3 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 13 | 1 | 2 | 2 | 1 | 5 | 0 | 2 | 10 | 20 | 685 | 0 | 4 | 0 | 752 | 91.09% |
| | X | 0 | 4 | 0 | 3 | 33 | 1 | 1 | 6 | 14 | 3 | 23 | 3 | 2 | 1 | 3 | 2 | 5 | 8 | 15 | 2 | 5 | 0 | 0 | 637 | 1 | 15 | 787 | 80.94% |
| | Y | 2 | 3 | 0 | 1 | 0 | 7 | 0 | 2 | 1 | 7 | 2 | 2 | 1 | 0 | 0 | 21 | 6 | 1 | 1 | 51 | 4 | 60 | 1 | 4 | 608 | 1 | 786 | 77.35% |
| | Z | 0 | 4 | 0 | 3 | 63 | 2 | 1 | 3 | 3 | 6 | 0 | 3 | 0 | 0 | 1 | 0 | 7 | 7 | 62 | 1 | 2 | 0 | 0 | 10 | 1 | 555 | 734 | 75.61% |
| Total | | 785 | 682 | 679 | 991 | 889 | 758 | 749 | 788 | 707 | 765 | 599 | 743 | 784 | 722 | 709 | 789 | 806 | 765 | 761 | 793 | 851 | 761 | 738 | 913 | 777 | 696 | 20000 | 79.24% |



**Letter * Predicted Value for Letter Crosstabulation(K=7, Binned)**

Count

| | | Predicted Value for Letter | | | | | | | | | | | | | | | | | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | | |
| Letter | A | 741 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 9 | 8 | 5 | 1 | 1 | 0 | 5 | 4 | 1 | 1 | 0 | 0 | 1 | 5 | 1 | 789 | 93.92% |
| | B | 2 | 566 | 0 | 21 | 6 | 15 | 3 | 3 | 2 | 1 | 3 | 1 | 3 | 3 | 4 | 1 | 12 | 51 | 18 | 5 | 6 | 5 | 2 | 22 | 7 | 4 | 766 | 73.89% |
| | C | 1 | 4 | 561 | 6 | 22 | 7 | 26 | 3 | 0 | 1 | 13 | 1 | 0 | 0 | 19 | 1 | 9 | 9 | 8 | 9 | 21 | 3 | 3 | 3 | 4 | 2 | 736 | 76.22% |
| | D | 0 | 26 | 2 | 663 | 2 | 2 | 2 | 12 | 1 | 3 | 4 | 0 | 2 | 14 | 22 | 5 | 8 | 14 | 5 | 4 | 5 | 0 | 0 | 4 | 4 | 1 | 805 | 82.36% |
| | E | 1 | 19 | 13 | 0 | 506 | 3 | 22 | 7 | 0 | 1 | 31 | 4 | 1 | 1 | 2 | 2 | 9 | 6 | 34 | 4 | 0 | 0 | 2 | 39 | 1 | 60 | 768 | 65.89% |
| | F | 1 | 9 | 2 | 9 | 2 | 593 | 0 | 2 | 9 | 4 | 1 | 2 | 0 | 8 | 0 | 43 | 1 | 6 | 7 | 55 | 4 | 4 | 3 | 3 | 6 | 1 | 775 | 76.52% |
| | G | 3 | 15 | 17 | 12 | 18 | 2 | 557 | 8 | 0 | 1 | 9 | 0 | 3 | 4 | 17 | 8 | 33 | 28 | 2 | 4 | 9 | 6 | 3 | 10 | 3 | 1 | 773 | 72.06% |
| | H | 3 | 31 | 1 | 32 | 5 | 4 | 7 | 475 | 25 | 0 | 19 | 3 | 3 | 17 | 24 | 12 | 6 | 27 | 6 | 1 | 9 | 0 | 1 | 17 | 3 | 3 | 734 | 64.71% |
| | I | 2 | 11 | 1 | 2 | 1 | 6 | 1 | 0 | 680 | 26 | 2 | 3 | 0 | 0 | 0 | 0 | 3 | 1 | 6 | 2 | 0 | 0 | 0 | 2 | 2 | 4 | 755 | 90.07% |
| | J | 2 | 8 | 0 | 8 | 1 | 12 | 2 | 3 | 54 | 598 | 1 | 6 | 0 | 6 | 2 | 2 | 5 | 6 | 12 | 1 | 2 | 0 | 0 | 9 | 5 | 2 | 747 | 80.05% |
| | K | 3 | 25 | 4 | 7 | 16 | 4 | 15 | 36 | 3 | 1 | 478 | 5 | 2 | 16 | 3 | 5 | 3 | 32 | 8 | 3 | 3 | 4 | 0 | 48 | 3 | 12 | 739 | 64.68% |
| | L | 28 | 1 | 0 | 1 | 6 | 1 | 5 | 2 | 4 | 1 | 1 | 674 | 0 | 1 | 0 | 3 | 4 | 11 | 3 | 1 | 1 | 0 | 0 | 10 | 3 | 0 | 761 | 88.57% |
| | M | 12 | 10 | 1 | 2 | 0 | 1 | 1 | 1 | 0 | 0 | 7 | 0 | 711 | 10 | 1 | 1 | 1 | 8 | 1 | 2 | 11 | 2 | 9 | 0 | 0 | 0 | 792 | 89.77% |
| | N | 7 | 9 | 0 | 34 | 0 | 0 | 4 | 15 | 2 | 1 | 2 | 0 | 13 | 642 | 16 | 2 | 1 | 7 | 0 | 0 | 10 | 5 | 11 | 0 | 2 | 0 | 783 | 81.99% |
| | O | 0 | 1 | 9 | 19 | 0 | 1 | 6 | 10 | 40 | 2 | 0 | 1 | 2 | 2 | 582 | 3 | 41 | 6 | 6 | 1 | 6 | 6 | 7 | 1 | 1 | 0 | 753 | 77.29% |
| | P | 1 | 15 | 0 | 23 | 1 | 75 | 8 | 4 | 2 | 3 | 2 | 4 | 0 | 3 | 3 | 600 | 5 | 6 | 8 | 11 | 0 | 2 | 1 | 4 | 22 | 0 | 803 | 74.72% |
| | Q | 1 | 9 | 2 | 14 | 5 | 0 | 14 | 4 | 4 | 1 | 3 | 1 | 1 | 1 | 33 | 4 | 656 | 3 | 11 | 1 | 3 | 2 | 2 | 3 | 4 | 1 | 783 | 83.78% |
| | R | 1 | 74 | 0 | 27 | 6 | 2 | 5 | 14 | 6 | 5 | 20 | 1 | 10 | 6 | 3 | 3 | 6 | 531 | 7 | 5 | 1 | 0 | 2 | 22 | 1 | 0 | 758 | 70.05% |
| | S | 3 | 41 | 1 | 8 | 30 | 10 | 5 | 1 | 4 | 7 | 0 | 0 | 3 | 3 | 3 | 4 | 3 | 13 | 504 | 5 | 4 | 1 | 0 | 14 | 3 | 78 | 748 | 67.38% |
| | T | 0 | 9 | 1 | 3 | 6 | 25 | 1 | 5 | 0 | 0 | 1 | 2 | 1 | 3 | 2 | 3 | 3 | 12 | 3 | 629 | 6 | 6 | 1 | 5 | 63 | 6 | 796 | 79.02% |
| | U | 4 | 0 | 5 | 5 | 0 | 4 | 1 | 2 | 1 | 1 | 0 | 0 | 7 | 7 | 12 | 1 | 5 | 1 | 0 | 3 | 739 | 7 | 3 | 1 | 4 | 0 | 813 | 90.90% |
| | V | 1 | 13 | 0 | 3 | 0 | 9 | 1 | 1 | 2 | 0 | 2 | 0 | 0 | 6 | 0 | 6 | 1 | 3 | 3 | 13 | 4 | 604 | 19 | 0 | 73 | 0 | 764 | 79.06% |
| | W | 0 | 8 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 15 | 2 | 3 | 0 | 0 | 5 | 1 | 0 | 9 | 15 | 688 | 0 | 3 | 0 | 752 | 91.49% |
| | X | 3 | 11 | 0 | 4 | 13 | 2 | 1 | 7 | 15 | 4 | 25 | 3 | 0 | 1 | 3 | 1 | 7 | 8 | 24 | 2 | 7 | 0 | 0 | 616 | 1 | 29 | 787 | 78.27% |
| | Y | 1 | 3 | 0 | 0 | 1 | 5 | 1 | 4 | 1 | 0 | 0 | 1 | 2 | 1 | 2 | 9 | 2 | 1 | 2 | 52 | 4 | 58 | 2 | 3 | 630 | 1 | 786 | 80.15% |
| | Z | 0 | 12 | 1 | 6 | 46 | 1 | 1 | 0 | 6 | 4 | 0 | 4 | 0 | 0 | 2 | 0 | 7 | 9 | 44 | 4 | 1 | 0 | 0 | 8 | 3 | 575 | 734 | 78.34% |
| Total | | 821 | 932 | 621 | 911 | 693 | 784 | 690 | 621 | 862 | 665 | 625 | 725 | 787 | 762 | 759 | 720 | 831 | 809 | 727 | 818 | 866 | 730 | 759 | 845 | 856 | 781 | 20000 | 78.89% |

IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

Question 2-c:
Ans:
Based on the above result, the result from the K-nearest neighbor seems have a more accurate result than the decision tree. However, The running time of K-nearest neighbor is a lot longer than the Decision. In my opinion, the accuracy measure may be a good idea for comparing the Training and Testing data for further verification. In the SPSS, there is not a lot of parameters setting available for K-nearest neighbor analysis. On the contrary, there is more adjustment available (Depth, Parents node and Child nodes etc.) on the decision tree algorithms. Also, I would prefer to use Decision Tree algorithms for the future analysis.

IS467 Name: Kai Chung, Ying
Homework#4 Email: kying@mail.depaul.edu
Date: 06-03-2017

Question3
a.

1) In k-means how are the cluster centers calculated?

Ans:
First of all, clusters are defined by their centers, the following is the steps to process calculation.

- First, the algorithm artitrarily choose K object as initial cluster center.
- And then assign each objects to most similar center.
- Then clusters update their cluster Centroids (i.e. Mean point)
- Then the objects will be assigned to the most similar center by the distance function
- Repeat step 3and 4 until NO change (=No better result) of centroids

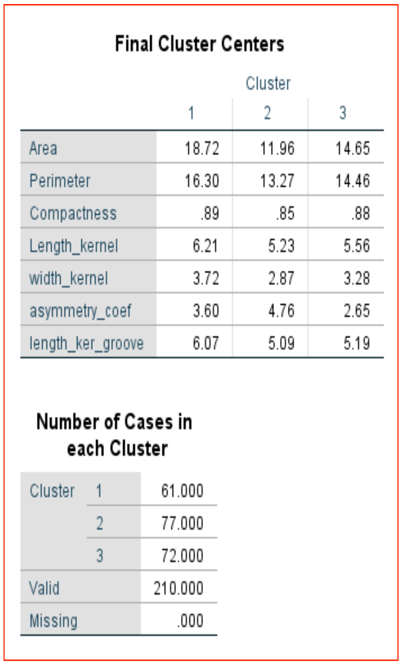
2) Name two similarity measures ( or distance functions) and what type of data you would use them for.

Ans:

- Minkowski Distance – which is good for two p-dimensional data object and q is a positive integer
- Manhattan distance – which is  good for 2-dimensional data, especially only have it (x, y) coordinates along the axis and q=1
- Euclidean distance – which is good for 2-dimensional data and with q=2.

3) Perform k-means clustering:
i & ii Report the final cluster centers and the number of elements in each clustering

k=3:



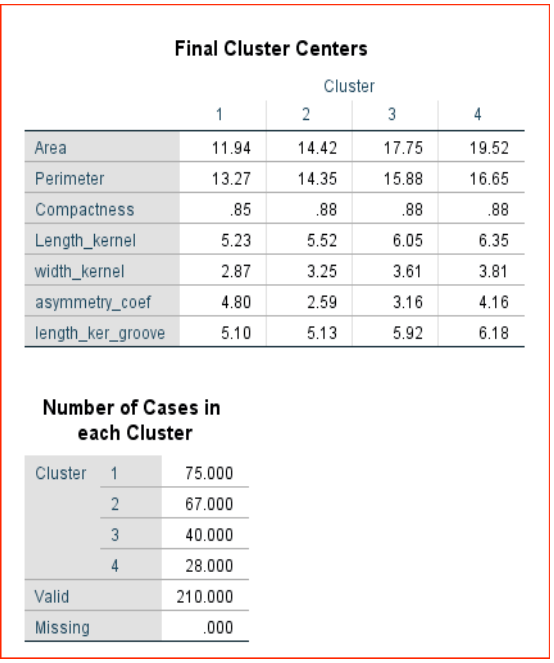
**Final Cluster Centers**

| | Cluster | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Area | 18.72 | 11.96 | 14.65 |
| Perimeter | 16.30 | 13.27 | 14.46 |
| Compactness | .89 | .85 | .88 |
| Length_kernel | 6.21 | 5.23 | 5.56 |
| width_kernel | 3.72 | 2.87 | 3.28 |
| asymmetry_coef | 3.60 | 4.76 | 2.65 |
| length_ker_groove | 6.07 | 5.09 | 5.19 |

**Number of Cases in each Cluster**

| | | |
|---|---|---|
| Cluster | 1 | 61.000 |
| | 2 | 77.000 |
| | 3 | 72.000 |
| Valid | | 210.000 |
| Missing | | .000 |

k=4:



**Final Cluster Centers**

| | Cluster | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Area | 11.94 | 14.42 | 17.75 | 19.52 |
| Perimeter | 13.27 | 14.35 | 15.88 | 16.65 |
| Compactness | .85 | .88 | .88 | .88 |
| Length_kernel | 5.23 | 5.52 | 6.05 | 6.35 |
| width_kernel | 2.87 | 3.25 | 3.61 | 3.81 |
| asymmetry_coef | 4.80 | 2.59 | 3.16 | 4.16 |
| length_ker_groove | 5.10 | 5.13 | 5.92 | 6.18 |

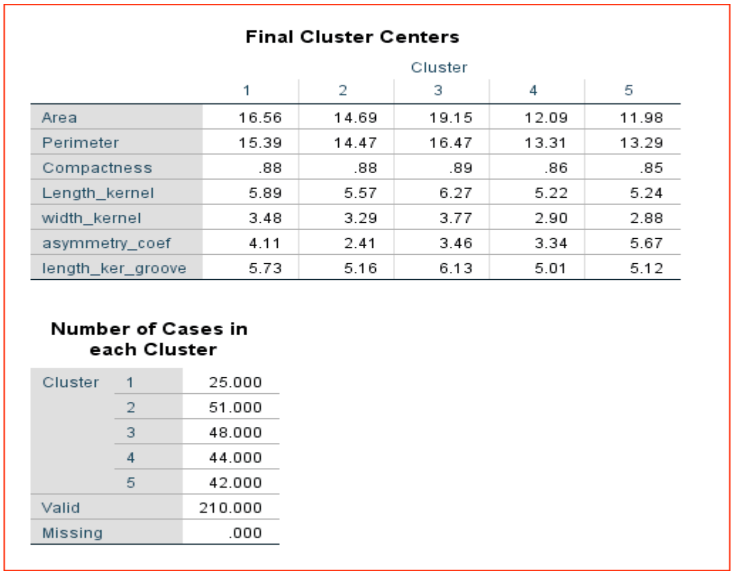
**Number of Cases in each Cluster**

| | | |
|---|---|---|
| Cluster | 1 | 75.000 |
| | 2 | 67.000 |
| | 3 | 40.000 |
| | 4 | 28.000 |
| Valid | | 210.000 |
| Missing | | .000 |

IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

k=5



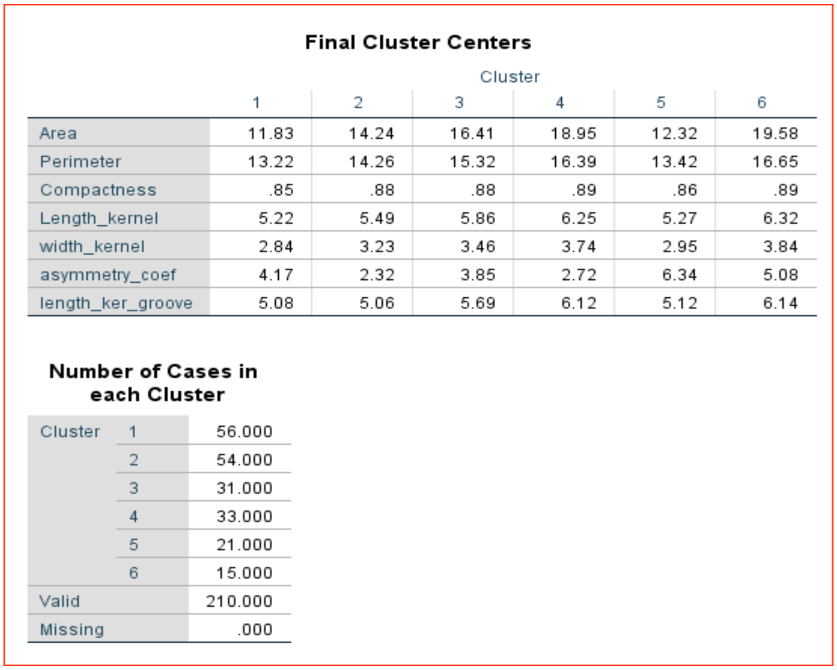
**Final Cluster Centers**

| | Cluster | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Area | 16.56 | 14.69 | 19.15 | 12.09 | 11.98 |
| Perimeter | 15.39 | 14.47 | 16.47 | 13.31 | 13.29 |
| Compactness | .88 | .88 | .89 | .86 | .85 |
| Length_kernel | 5.89 | 5.57 | 6.27 | 5.22 | 5.24 |
| width_kernel | 3.48 | 3.29 | 3.77 | 2.90 | 2.88 |
| asymmetry_coef | 4.11 | 2.41 | 3.46 | 3.34 | 5.67 |
| length_ker_groove | 5.73 | 5.16 | 6.13 | 5.01 | 5.12 |

**Number of Cases in each Cluster**

| | | |
|---|---|---|
| Cluster | 1 | 25.000 |
| | 2 | 51.000 |
| | 3 | 48.000 |
| | 4 | 44.000 |
| | 5 | 42.000 |
| Valid | | 210.000 |
| Missing | | .000 |

k=6



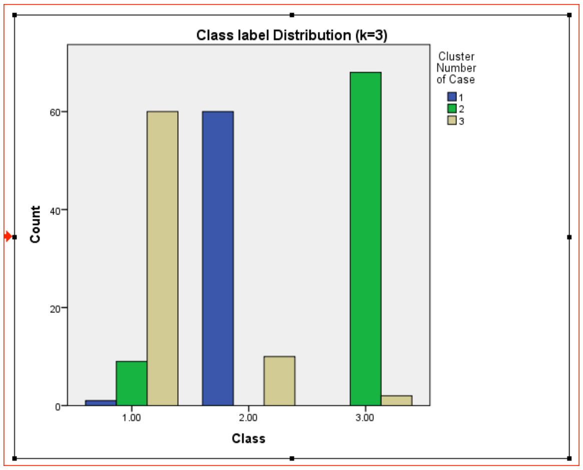
**Final Cluster Centers**

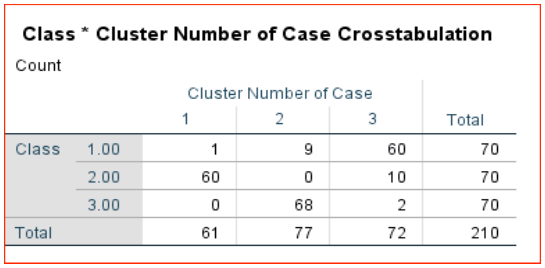
| | Cluster | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Area | 11.83 | 14.24 | 16.41 | 18.95 | 12.32 | 19.58 |
| Perimeter | 13.22 | 14.26 | 15.32 | 16.39 | 13.42 | 16.65 |
| Compactness | .85 | .88 | .88 | .89 | .86 | .89 |
| Length_kernel | 5.22 | 5.49 | 5.86 | 6.25 | 5.27 | 6.32 |
| width_kernel | 2.84 | 3.23 | 3.46 | 3.74 | 2.95 | 3.84 |
| asymmetry_coef | 4.17 | 2.32 | 3.85 | 2.72 | 6.34 | 5.08 |
| length_ker_groove | 5.08 | 5.06 | 5.69 | 6.12 | 5.12 | 6.14 |

**Number of Cases in each Cluster**

| | | |
|---|---|---|
| Cluster | 1 | 56.000 |
| | 2 | 54.000 |
| | 3 | 31.000 |
| | 4 | 33.000 |
| | 5 | 21.000 |
| | 6 | 15.000 |
| Valid | | 210.000 |
| Missing | | .000 |

iii.Report the class distribution within each cluster
(use crosstab between labels and cluster membership)
k=3



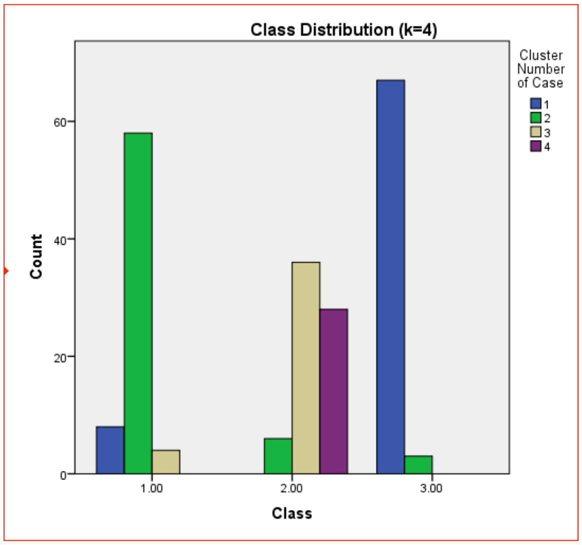


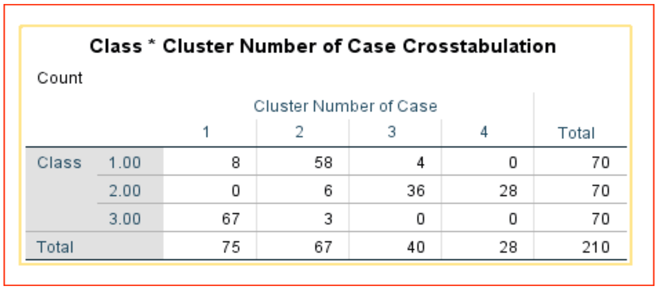
**Class * Cluster Number of Case Crosstabulation**

Count

| | | Cluster Number of Case | | | Total |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | |
| Class | 1.00 | 1 | 9 | 60 | 70 |
| | 2.00 | 60 | 0 | 10 | 70 |
| | 3.00 | 0 | 68 | 2 | 70 |
| Total | | 61 | 77 | 72 | 210 |

k=4





**Class * Cluster Number of Case Crosstabulation**

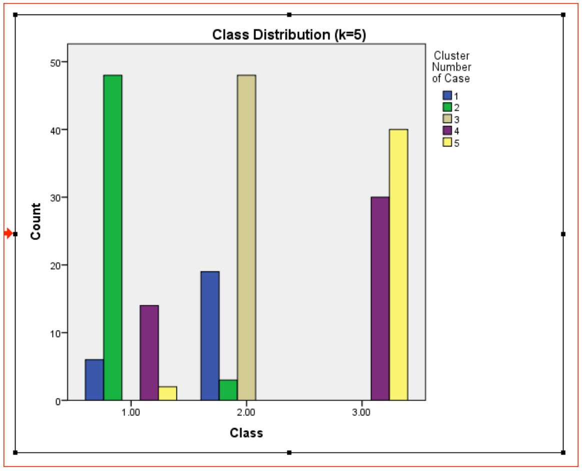
Count

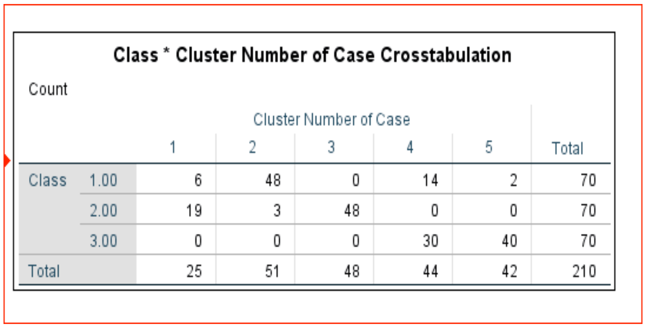
| | | Cluster Number of Case | | | | Total |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| Class | 1.00 | 8 | 58 | 4 | 0 | 70 |
| | 2.00 | 0 | 6 | 36 | 28 | 70 |
| | 3.00 | 67 | 3 | 0 | 0 | 70 |
| Total | | 75 | 67 | 40 | 28 | 210 |

IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

K=5



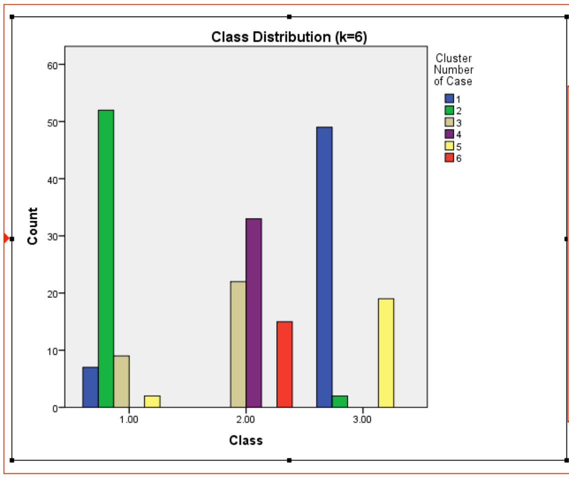


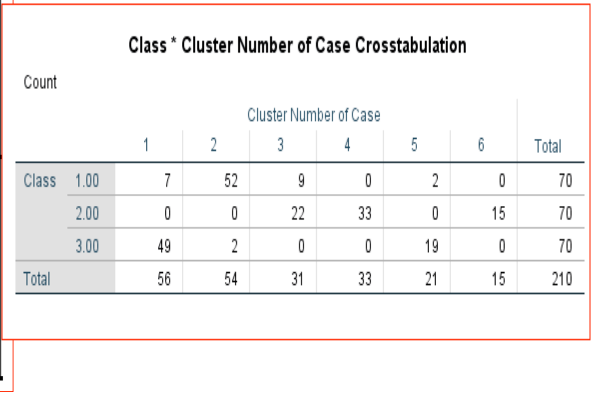
Class * Cluster Number of Case Crosstabulation

Count

| | | Cluster Number of Case | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | |
| Class | 1.00 | 6 | 48 | 0 | 14 | 2 | 70 |
| | 2.00 | 19 | 3 | 48 | 0 | 0 | 70 |
| | 3.00 | 0 | 0 | 0 | 30 | 40 | 70 |
| Total | | 25 | 51 | 48 | 44 | 42 | 210 |

K=6





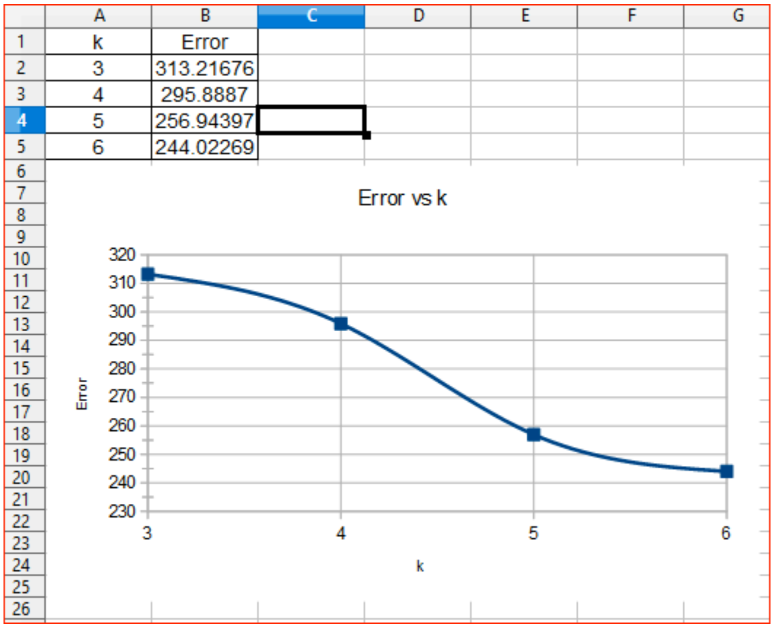
Class * Cluster Number of Case Crosstabulation

Count

| | | Cluster Number of Case | | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | |
| Class | 1.00 | 7 | 52 | 9 | 0 | 2 | 0 | 70 |
| | 2.00 | 0 | 0 | 22 | 33 | 0 | 15 | 70 |
| | 3.00 | 49 | 2 | 0 | 0 | 19 | 0 | 70 |
| Total | | 56 | 54 | 31 | 33 | 21 | 15 | 210 |

IS467 Name: Kai Chung, Ying
Homework#4 Email: kying@mail.depaul.edu
Date: 06-03-2017

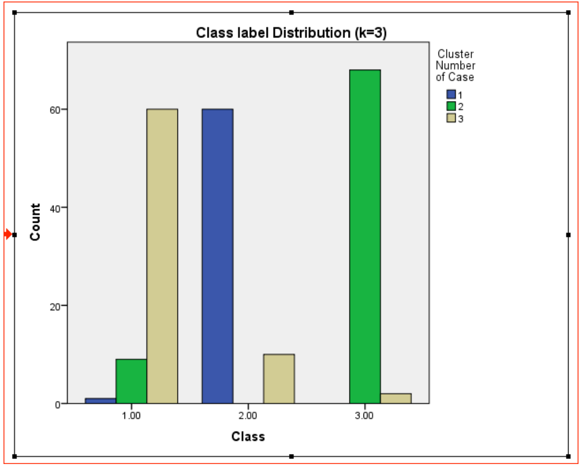
Question3a
Part4 )

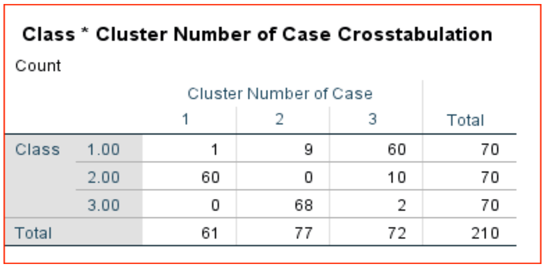


| k | Error |
|---|---|
| 3 | 313.21676 |
| 4 | 295.8887 |
| 5 | 256.94397 |
| 6 | 244.02269 |

Based on the above , I would choose the clustering with k=5 because the curve seems touch the knee after the k=5 section. In other words, there is not much improvements on the accuracy even increasing the number of clusters to the

Question3a
Part 5:
Ans: According to the part 3a iii), I would choose clusters with k=3 shown below. The reason is that the result most likely matches to the pattern of class labels originally provided from the Table even though their labels' number NOT named the same. Although the clustering is not perfectly classified, they majority of clusters matches to the classes labels, especially the Cluster #2 (Green bar) matches to the class label #3 more than 60 counts. So that I would conclude that k=3 has the best clustering result.





**Class * Cluster Number of Case Crosstabulation**

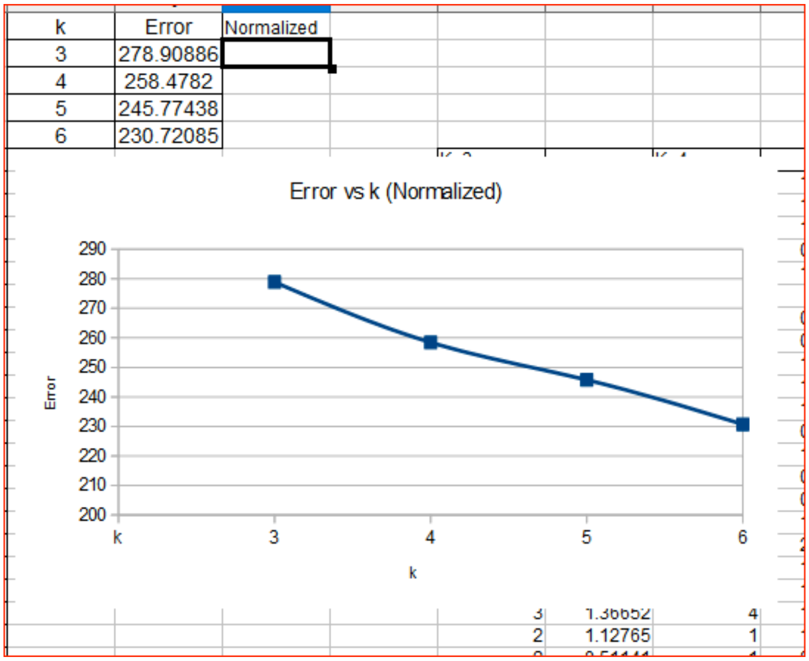
Count

| | | Cluster Number of Case | | | Total |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | |
| Class | 1.00 | 1 | 9 | 60 | 70 |
| | 2.00 | 60 | 0 | 10 | 70 |
| | 3.00 | 0 | 68 | 2 | 70 |
| Total | | 61 | 77 | 72 | 210 |

IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

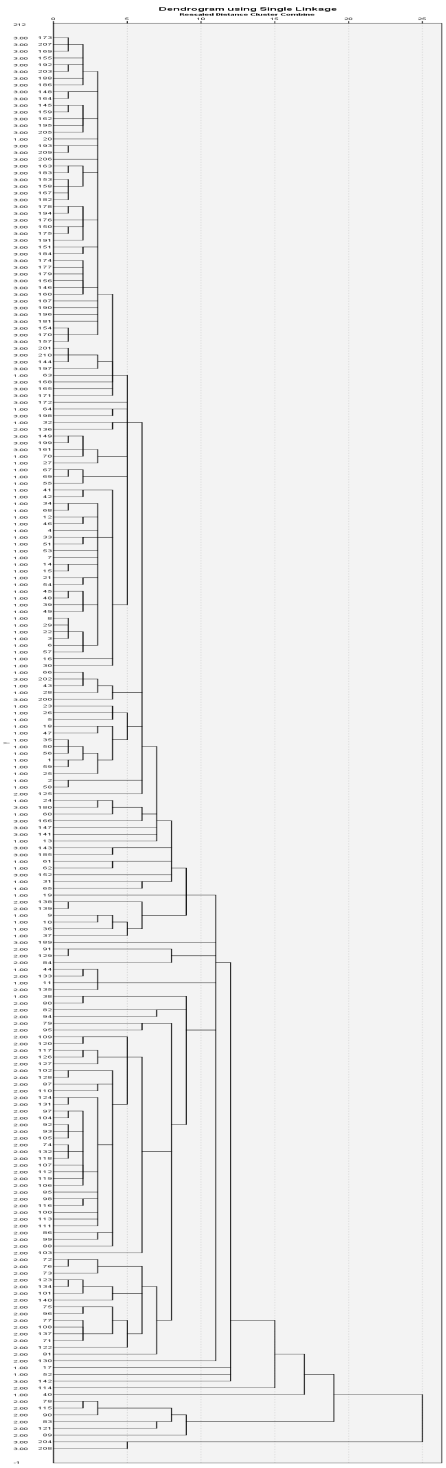
Question 3a
Part 6:
Based on the following plot for normalized data, I would say that the normalized data gives out a more accurate result. When we oberve the plot, we kind of  NOT able to see the knee or elbow section which means that more clusters (k >7) might be able to add for analysis to increase the accuracy. In part 4, the cluster's knee is at about  k=5 with error=256. In this result, the error probably lower than 230 with k>6. The reason behind that is probably some of the  attributes (e.g. Area, Perimeter) are too large comparing the other attributes in the dataset. And The distance function is sensitive to the data variables with very wide range because the equation involves square of the subtract values.

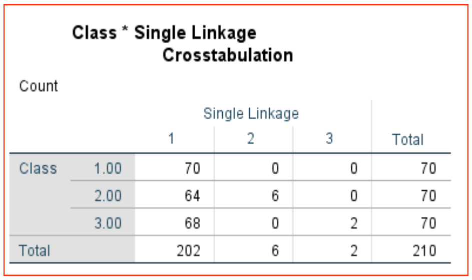


| k | Error | Normalized |
|---|---|---|
| 3 | 278.90886 | |
| 4 | 258.4782 | |
| 5 | 245.77438 | |
| 6 | 230.72085 | |

| | | |
|---|---|---|
| 3 | 1.36652 | 4 |
| 2 | 1.12765 | 1 |

Question3b - 1: Single Linkage algorithms and Reports

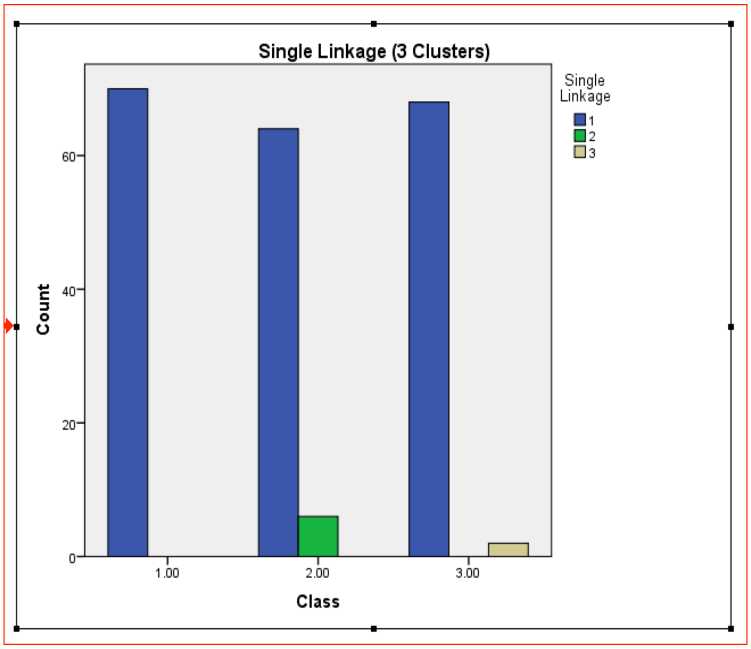




**Class * Single Linkage Crosstabulation**

Count

| | | Single Linkage | | | Total |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | |
| Class | 1.00 | 70 | 0 | 0 | 70 |
| | 2.00 | 64 | 6 | 0 | 70 |
| | 3.00 | 68 | 0 | 2 | 70 |
| Total | | 202 | 6 | 2 | 210 |

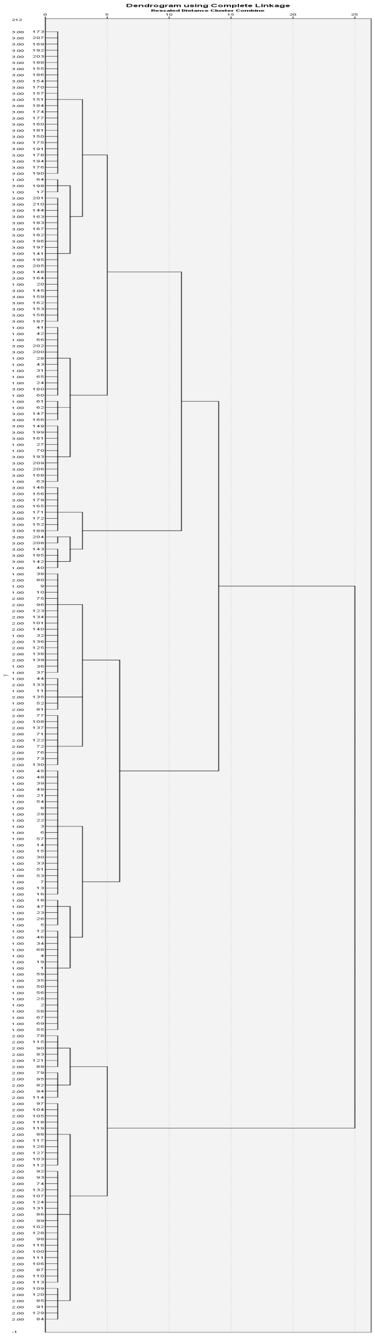


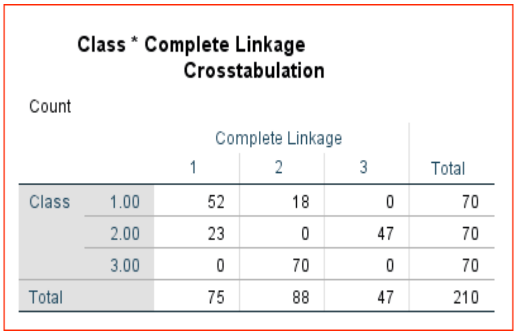
IS467<br>Homework#4<br>Date: 06-03-2017

Name: Kai Chung, Ying<br>Email: kying@mail.depaul.edu

Question3b - 2: Complete Linkage algorithms and Reports

Based on the the graph and tables, we can observe that the Compete Linkage Method's result is a lot more accurate than the Single Linkage Method. The Single Link 's result basically only came up 1 cluster(Blue color bar) , but it is supposed to have 3 clusters. On the contrary, the Complete Linkage came up 3 clusters and the predicted values is quite distinguishable between clusters.

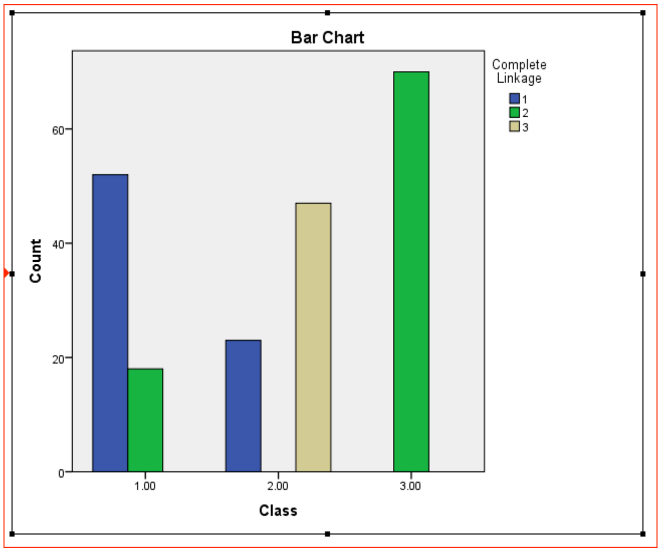

Dendrogram using Complete Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25



**Class * Complete Linkage Crosstabulation**

Count

| | | Complete Linkage | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | Total |
| Class | 1.00 | 52 | 18 | 0 | 70 |
| | 2.00 | 23 | 0 | 47 | 70 |
| | 3.00 | 0 | 70 | 0 | 70 |
| Total | | 75 | 88 | 47 | 210 |



IS467
Homework#4
Date: 06-03-2017

Name: Kai Chung, Ying
Email: kying@mail.depaul.edu

## Question 3c: Summary

Clustering analysis is the method we have done on the question 3. Clustering analysis is grouping a set of objects together based on their characteristic / similarity.  Generally speaking, the objects within a group (cluster) would have the most similar characteristic comparing the objects outside group. The way to group the objects together by using different distance functions (e.g. Euclidean) to calculate the distance between the reference point and the object and to partition objects into a group.

Two major clustering methods which are K-means and Hierarchical analysis were carried out above. K-means analysis is designed to partition the objects/observations into groups (clusters) with the closest cluster centroid (i.e. mean). K-means runs n times iterations to measure the distance of the objects within group and reassign the objects to their nearest cluster centroid. The iterations process normally ends when the clusters reaches the best performance (the distance between objects and centroid within group are shortest). Certainly, there are parameters (e.g. k is the number of clusters) that users have to pre-set before the algorithm start the iteration process. The clusters results could be illustrated by scatter plots, cross-tab matrix and Error graph as shown above. In this particular exercise, the Error graph is used to decide the performance of the clustering analysis. The graph shows the trend of errors changes by increasing the number of clusters. Technically, the more clusters assigned is supposed to lead to better result (less error). So that all we need to do is to observe the graph and find the k (number of cluster) section that starts showing less error improvements. Then that k-means cluster probably the best result of overall analysis.

Hierarchical clustering analysis uses another way to group the similar objects to clusters. Certainly, pre-set parameters (e.g. number of clusters) are necessary prior to the analysis. In this exercise, we have used single linkage and complete linkage algorithms. In the beginning the of the process, the objects are in their clusters of their own, and then the clusters merges by measuring the distance. The clusters combine based on the shortest distance. However, there are different way to define the shortest distance for these 2 methods. The shortest length of single linkage is the shortest distance between 2 members from the clusters; The shortest length of complete linkage is the farthest distance between 2 members from the clusters. With this method, we are not able to check the accuracy. We could just check the dendogram based on our desired clusters level and class distributions to decide if we would keep the result. For example, if we would like to reach 3 clusters, the result however came up only 2 clusters cases, then we might want to reset our parameters to give another try.

During the K-means clustering analysis, I experienced a very important step to lower the error on the given k (number of clusters) values. Since we are calculated the distance between objects / clusters. So that some skewed numerical attributes values (Too large or small) might influence the distance measures. So that we probably need to normalize the data prior to step in analysis. This would lead us to get the better result. Also, we have to convert the category variables to numerical in order to calculating the distance for clustering.