Alumno: Caio Fernandes Moreno

Fecha: 11 de Enero de 2015

El primer paso es cargar los datos quitando las columnas Obs e Species.

**data** peces (drop=Obs Species);

input Obs Weight Length1 Length2 Length3 HeightPCT WidthPCT Species;

datalines;

1 242.0 23.2 25.4 30.0 38.4 13.4 1

2 290.0 24.0 26.3 31.2 40.0 13.8 1

3 340.0 23.9 26.5 31.1 39.8 15.1 1

4 363.0 26.3 29.0 33.5 38.0 13.3 1

5 430.0 26.5 29.0 34.0 36.6 15.1 1

6 450.0 26.8 29.7 34.7 39.2 14.2 1

7 500.0 26.8 29.7 34.5 41.1 15.3 1

8 390.0 27.6 30.0 35.0 36.2 13.4 1

9 450.0 27.6 30.0 35.1 39.9 13.8 1

10 500.0 28.5 30.7 36.2 39.3 13.7 1

11 475.0 28.4 31.0 36.2 39.4 14.1 1

12 500.0 28.7 31.0 36.2 39.7 13.3 1

13 500.0 29.1 31.5 36.4 37.8 12.0 1

14 600.0 29.4 32.0 37.2 40.2 13.9 1

15 600.0 29.4 32.0 37.2 41.5 15.0 1

16 700.0 30.4 33.0 38.3 38.8 13.8 1

17 700.0 30.4 33.0 38.5 38.8 13.5 1

18 610.0 30.9 33.5 38.6 40.5 13.3 1

19 650.0 31.0 33.5 38.7 37.4 14.8 1

20 575.0 31.3 34.0 39.5 38.3 14.1 1

21 685.0 31.4 34.0 39.2 40.8 13.7 1

22 620.0 31.5 34.5 39.7 39.1 13.3 1

23 680.0 31.8 35.0 40.6 38.1 15.1 1

24 700.0 31.9 35.0 40.5 40.1 13.8 1

25 725.0 31.8 35.0 40.9 40.0 14.8 1

26 720.0 32.0 35.0 40.6 40.3 15.0 1

27 714.0 32.7 36.0 41.5 39.8 14.1 1

28 850.0 32.8 36.0 41.6 40.6 14.9 1

29 1000.0 33.5 37.0 42.6 44.5 15.5 1

30 920.0 35.0 38.5 44.1 40.9 14.3 1

31 955.0 35.0 38.5 44.0 41.1 14.3 1

32 925.0 36.2 39.5 45.3 41.4 14.9 1

33 975.0 37.4 41.0 45.9 40.6 14.7 1

34 950.0 38.0 41.0 46.5 37.9 13.7 1

35 270.0 23.6 26.0 28.7 29.2 14.8 2

36 270.0 24.1 26.5 29.3 27.8 14.5 2

37 306.0 25.6 28.0 30.8 28.5 15.2 2

38 540.0 28.5 31.0 34.0 31.6 19.3 2

39 800.0 33.7 36.4 39.6 29.7 16.6 2

40 1000.0 37.3 40.0 43.5 28.4 15.0 2

41 40.0 12.9 14.1 16.2 25.6 14.0 3

42 69.0 16.5 18.2 20.3 26.1 13.9 3

43 78.0 17.5 18.8 21.2 26.3 13.7 3

44 87.0 18.2 19.8 22.2 25.3 14.3 3

45 120.0 18.6 20.0 22.2 28.0 16.1 3

46 0.0 19.0 20.5 22.8 28.4 14.7 3

47 110.0 19.1 20.8 23.1 26.7 14.7 3

48 120.0 19.4 21.0 23.7 25.8 13.9 3

49 150.0 20.4 22.0 24.7 23.5 15.2 3

50 145.0 20.5 22.0 24.3 27.3 14.6 3

51 160.0 20.5 22.5 25.3 27.8 15.1 3

52 140.0 21.0 22.5 25.0 26.2 13.3 3

53 160.0 21.1 22.5 25.0 25.6 15.2 3

54 169.0 22.0 24.0 27.2 27.7 14.1 3

55 161.0 22.0 23.4 26.7 25.9 13.6 3

56 200.0 22.1 23.5 26.8 27.6 15.4 3

57 180.0 23.6 25.2 27.9 25.4 14.0 3

58 290.0 24.0 26.0 29.2 30.4 15.4 3

59 272.0 25.0 27.0 30.6 28.0 15.6 3

60 390.0 29.5 31.7 35.0 27.1 15.3 3

61 55.0 13.5 14.7 16.5 41.5 14.1 4

62 60.0 14.3 15.5 17.4 37.8 13.3 4

63 90.0 16.3 17.7 19.8 37.4 13.5 4

64 120.0 17.5 19.0 21.3 39.4 13.7 4

65 150.0 18.4 20.0 22.4 39.7 14.7 4

66 140.0 19.0 20.7 23.2 36.8 14.2 4

67 170.0 19.0 20.7 23.2 40.5 14.7 4

68 145.0 19.8 21.5 24.1 40.4 13.1 4

69 200.0 21.2 23.0 25.8 40.1 14.2 4

70 273.0 23.0 25.0 28.0 39.6 14.8 4

71 300.0 24.0 26.0 29.0 39.2 14.6 4

72 6.7 9.3 9.8 10.8 16.1 9.7 5

73 7.5 10.0 10.5 11.6 17.0 10.0 5

74 7.0 10.1 10.6 11.6 14.9 9.9 5

75 9.7 10.4 11.0 12.0 18.3 11.5 5

76 9.8 10.7 11.2 12.4 16.8 10.3 5

77 8.7 10.8 11.3 12.6 15.7 10.2 5

78 10.0 11.3 11.8 13.1 16.9 9.8 5

79 9.9 11.3 11.8 13.1 16.9 8.9 5

80 9.8 11.4 12.0 13.2 16.7 8.7 5

81 12.2 11.5 12.2 13.4 15.6 10.4 5

82 13.4 11.7 12.4 13.5 18.0 9.4 5

83 12.2 12.1 13.0 13.8 16.5 9.1 5

84 19.7 13.2 14.3 15.2 18.9 13.6 5

85 19.9 13.8 15.0 16.2 18.1 11.6 5

86 200.0 30.0 32.3 34.8 16.0 9.7 6

87 300.0 31.7 34.0 37.8 15.1 11.0 6

88 300.0 32.7 35.0 38.8 15.3 11.3 6

89 300.0 34.8 37.3 39.8 15.8 10.1 6

90 430.0 35.5 38.0 40.5 18.0 11.3 6

91 345.0 36.0 38.5 41.0 15.6 9.7 6

92 456.0 40.0 42.5 45.5 16.0 9.5 6

93 510.0 40.0 42.5 45.5 15.0 9.8 6

94 540.0 40.1 43.0 45.8 17.0 11.2 6

95 500.0 42.0 45.0 48.0 14.5 10.2 6

96 567.0 43.2 46.0 48.7 16.0 10.0 6

97 770.0 44.8 48.0 51.2 15.0 10.5 6

98 950.0 48.3 51.7 55.1 16.2 11.2 6

99 1250.0 52.0 56.0 59.7 17.9 11.7 6

100 1600.0 56.0 60.0 64.0 15.0 9.6 6

101 1550.0 56.0 60.0 64.0 15.0 9.6 6

102 1650.0 59.0 63.4 68.0 15.9 11.0 6

103 5.9 7.5 8.4 8.8 24.0 16.0 7

104 32.0 12.5 13.7 14.7 24.0 13.6 7

105 40.0 13.8 15.0 16.0 23.9 15.2 7

106 51.5 15.0 16.2 17.2 26.7 15.3 7

107 70.0 15.7 17.4 18.5 24.8 15.9 7

108 100.0 16.2 18.0 19.2 27.2 17.3 7

109 78.0 16.8 18.7 19.4 26.8 16.1 7

110 80.0 17.2 19.0 20.2 27.9 15.1 7

111 85.0 17.8 19.6 20.8 24.7 14.6 7

112 85.0 18.2 20.0 21.0 24.2 13.2 7

113 110.0 19.0 21.0 22.5 25.3 15.8 7

114 115.0 19.0 21.0 22.5 26.3 14.7 7

115 125.0 19.0 21.0 22.5 25.3 16.3 7

116 130.0 19.3 21.3 22.8 28.0 15.5 7

117 120.0 20.0 22.0 23.5 26.0 14.5 7

118 120.0 20.0 22.0 23.5 24.0 15.0 7

119 130.0 20.0 22.0 23.5 26.0 15.0 7

120 135.0 20.0 22.0 23.5 25.0 15.0 7

121 110.0 20.0 22.0 23.5 23.5 17.0 7

122 130.0 20.5 22.5 24.0 24.4 15.1 7

123 150.0 20.5 22.5 24.0 28.3 15.1 7

124 145.0 20.7 22.7 24.2 24.6 15.0 7

125 150.0 21.0 23.0 24.5 21.3 14.8 7

126 170.0 21.5 23.5 25.0 25.1 14.9 7

127 225.0 22.0 24.0 25.5 28.6 14.6 7

128 145.0 22.0 24.0 25.5 25.0 15.0 7

129 188.0 22.6 24.6 26.2 25.7 15.9 7

130 180.0 23.0 25.0 26.5 24.3 13.9 7

131 197.0 23.5 25.6 27.0 24.3 15.7 7

132 218.0 25.0 26.5 28.0 25.6 14.8 7

133 300.0 25.2 27.3 28.7 29.0 17.9 7

134 260.0 25.4 27.5 28.9 24.8 15.0 7

135 265.0 25.4 27.5 28.9 24.4 15.0 7

136 250.0 25.4 27.5 28.9 25.2 15.8 7

137 250.0 25.9 28.0 29.4 26.6 14.3 7

138 300.0 26.9 28.7 30.1 25.2 15.4 7

139 320.0 27.8 30.0 31.6 24.1 15.1 7

140 514.0 30.5 32.8 34.0 29.5 17.7 7

141 556.0 32.0 34.5 36.5 28.1 17.5 7

142 840.0 32.5 35.0 37.3 30.8 20.9 7

143 685.0 34.0 36.5 39.0 27.9 17.6 7

144 700.0 34.0 36.0 38.3 27.7 17.6 7

145 700.0 34.5 37.0 39.4 27.5 15.9 7

146 690.0 34.6 37.0 39.3 26.9 16.2 7

147 900.0 36.5 39.0 41.4 26.9 18.1 7

148 650.0 36.5 39.0 41.4 26.9 14.5 7

149 820.0 36.6 39.0 41.3 30.1 17.8 7

150 850.0 36.9 40.0 42.3 28.2 16.8 7

151 900.0 37.0 40.0 42.5 27.6 17.0 7

152 1015.0 37.0 40.0 42.4 29.2 17.6 7

153 820.0 37.1 40.0 42.5 26.2 15.6 7

154 1100.0 39.0 42.0 44.6 28.7 15.4 7

155 1000.0 39.8 43.0 45.2 26.4 16.1 7

156 1100.0 40.1 43.0 45.5 27.5 16.3 7

157 1000.0 40.2 43.5 46.0 27.4 17.7 7

158 1000.0 41.1 44.0 46.6 26.8 16.3 7

;

**proc** **print**;

**run**;

Como las variables numéricas tienen distintas escalas es necesario normalizar los datos utilizando el procedimiento stdize.

**proc** **stdize** data=peces out=pecesnorm;

**run**;

Ahora probamos con distintos metodos como average, centroid y ward.

**proc** **cluster** data=pecesnorm method=average pseudo RSQUARE ccc outtree=pecesA

print=**10** plots=den(VERTICAL);

**run**;

**proc** **cluster** data=pecesnorm method=centroid pseudo ccc RSQUARE

outtree=pecesC

print=**10** plots=den(VERTICAL);

**run**;

**proc** **cluster** data=pecesnorm method=ward pseudo ccc RSQUARE

outtree=pecesW

print=**10** plots=den(VERTICAL);

**run**;

Quedamos con 4 a 5 grupos (clusters).

Resultados:

|  |
| --- |
|  |

The CLUSTER Procedure

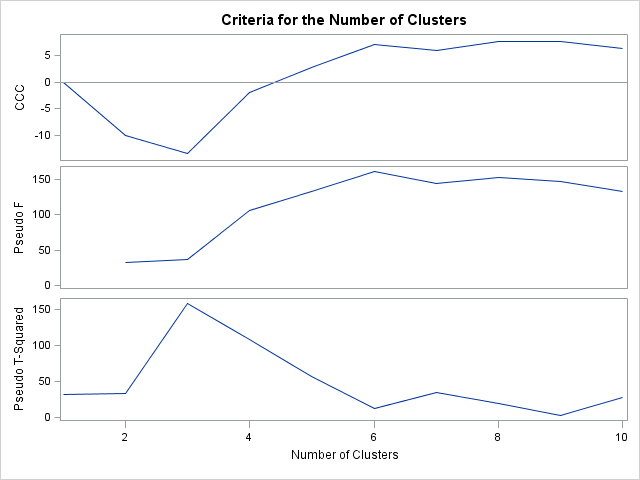
**Average** Linkage Cluster Analysis

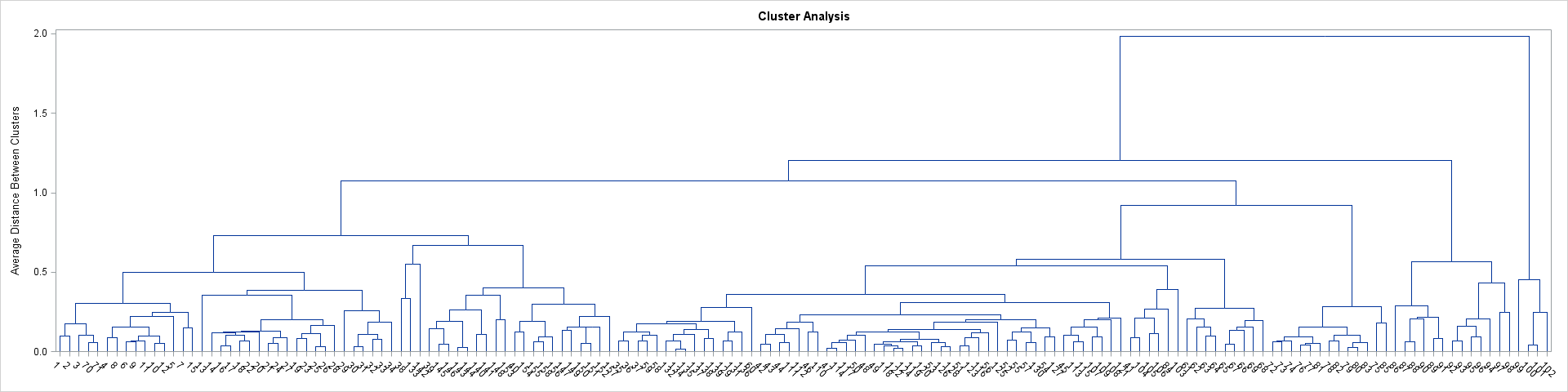
|  | | | | |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

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

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

| **Cluster History** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number of Clusters** | **Clusters Joined** | | **Freq** | **Semipartial R-Square** | **R-Square** | **Approximate Expected R-Square** | **Cubic Clustering Criterion** | **Pseudo F Statistic** | **Pseudo t-Squared** | **Norm RMS Distance** | **Tie** |
| **10** | **CL17** | **CL15** | 60 | 0.0150 | .890 | .845 | 6.39 | 133 | 28.3 | 0.5414 |  |
| **9** | **CL20** | **OB142** | 3 | 0.0023 | .887 | .833 | 7.59 | 147 | 3.2 | 0.5479 |  |
| **8** | **CL24** | **CL13** | 13 | 0.0105 | .877 | .818 | 7.68 | 153 | 20.1 | 0.5681 |  |
| **7** | **CL10** | **CL27** | 69 | 0.0252 | .852 | .800 | 6.06 | 145 | 34.9 | 0.5797 |  |
| **6** | **CL9** | **CL14** | 23 | 0.0104 | .841 | .777 | 7.13 | 161 | 12.7 | 0.6693 |  |
| **5** | **CL11** | **CL6** | 59 | 0.0644 | .777 | .746 | 2.80 | 133 | 57.4 | 0.7306 |  |
| **4** | **CL7** | **CL25** | 82 | 0.1039 | .673 | .704 | -1.9 | 106 | 109 | 0.9191 |  |
| **3** | **CL5** | **CL4** | 141 | 0.3526 | .321 | .635 | -13 | 36.6 | 159 | 1.0722 |  |
| **2** | **CL3** | **CL8** | 154 | 0.1483 | .172 | .492 | -10 | 32.5 | 33.3 | 1.2018 |  |
| **1** | **CL2** | **CL12** | 158 | 0.1722 | .000 | .000 | 0.00 | . | 32.5 | 1.9835 |  |



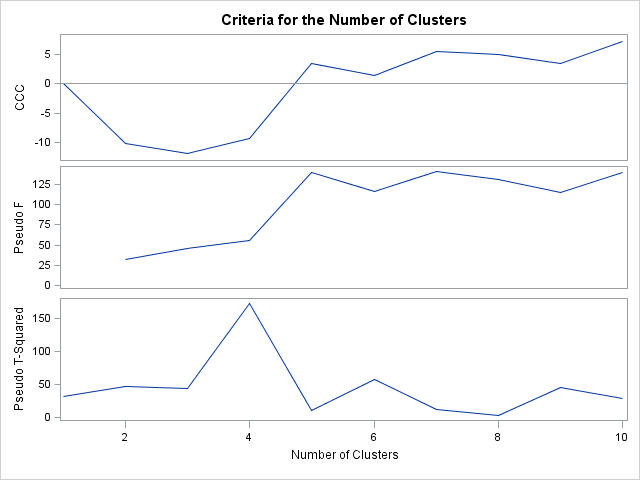


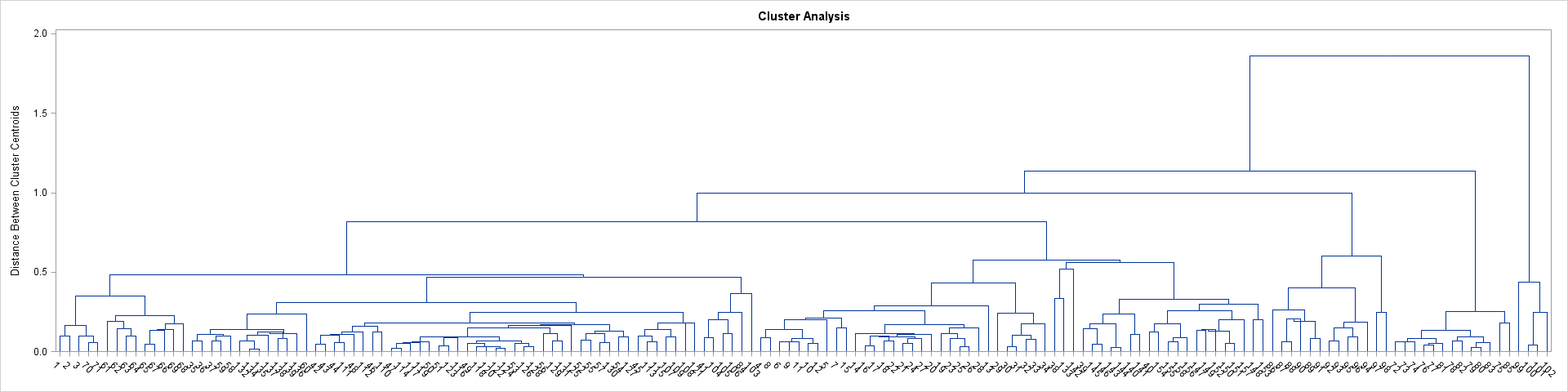
|  |
| --- |
|  |

The CLUSTER Procedure

**Centroid** Hierarchical Cluster Analysis

| **Cluster History** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number of Clusters** | **Clusters Joined** | | **Freq** | **Semipartial R-Square** | **R-Square** | **Approximate Expected R-Square** | **Cubic Clustering Criterion** | **Pseudo F Statistic** | **Pseudo t-Squared** | **Norm Centroid Distance** | **Tie** |
| **10** | **CL18** | **CL14** | 60 | 0.0150 | .895 | .845 | 7.27 | 140 | 28.3 | 0.4665 |  |
| **9** | **CL15** | **CL10** | 74 | 0.0338 | .861 | .833 | 3.55 | 115 | 45.1 | 0.4833 |  |
| **8** | **CL16** | **OB142** | 3 | 0.0023 | .859 | .818 | 4.98 | 130 | 3.2 | 0.5215 |  |
| **7** | **CL8** | **CL17** | 23 | 0.0104 | .848 | .800 | 5.60 | 141 | 12.7 | 0.5594 |  |
| **6** | **CL12** | **CL7** | 54 | 0.0560 | .792 | .777 | 1.52 | 116 | 57.5 | 0.577 |  |
| **5** | **CL13** | **CL28** | 13 | 0.0079 | .784 | .746 | 3.55 | 139 | 10.3 | 0.6038 |  |
| **4** | **CL9** | **CL6** | 128 | 0.2665 | .518 | .704 | -9.3 | 55.2 | 173 | 0.8186 |  |
| **3** | **CL4** | **CL5** | 141 | 0.1500 | .368 | .635 | -12 | 45.1 | 43.7 | 0.9988 |  |
| **2** | **CL3** | **CL24** | 154 | 0.1958 | .172 | .492 | -10 | 32.5 | 47.3 | 1.1363 |  |
| **1** | **CL2** | **CL11** | 158 | 0.1722 | .000 | .000 | 0.00 | . | 32.5 | 1.8621 |  |



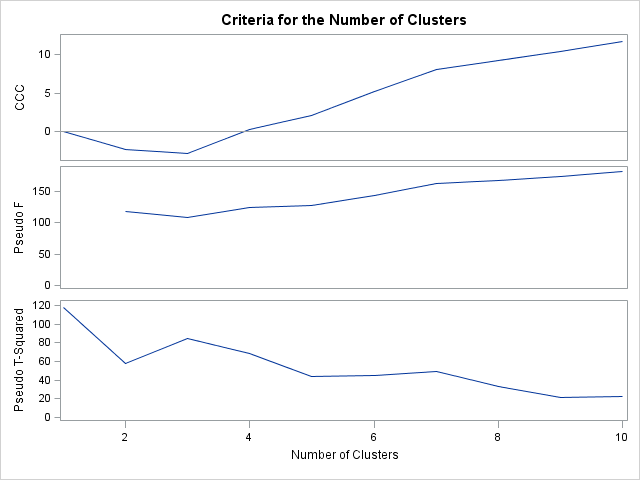


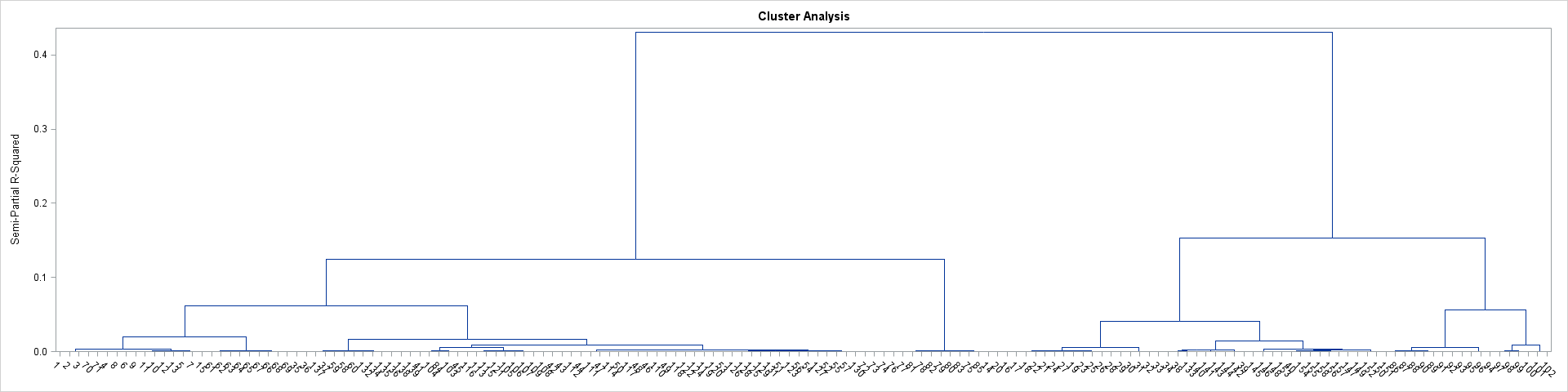
|  |
| --- |
|  |

The CLUSTER Procedure

**Ward's** Minimum Variance Cluster Analysis

| **Cluster History** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number of Clusters** | **Clusters Joined** | | **Freq** | **Semipartial R-Square** | **R-Square** | **Approximate Expected R-Square** | **Cubic Clustering Criterion** | **Pseudo F Statistic** | **Pseudo t-Squared** | **Tie** |
| **10** | **CL14** | **CL17** | 47 | 0.0091 | .917 | .845 | 11.7 | 182 | 22.9 |  |
| **9** | **CL18** | **CL15** | 23 | 0.0139 | .903 | .833 | 10.5 | 174 | 21.2 |  |
| **8** | **CL31** | **CL10** | 60 | 0.0165 | .887 | .818 | 9.28 | 168 | 32.8 |  |
| **7** | **CL16** | **CL23** | 25 | 0.0203 | .866 | .800 | 8.14 | 163 | 49.5 |  |
| **6** | **CL12** | **CL9** | 43 | 0.0407 | .826 | .777 | 5.16 | 144 | 45.5 |  |
| **5** | **CL13** | **CL11** | 17 | 0.0558 | .770 | .746 | 2.12 | 128 | 43.9 |  |
| **4** | **CL7** | **CL8** | 85 | 0.0620 | .708 | .704 | 0.24 | 124 | 68.3 |  |
| **3** | **CL4** | **CL24** | 98 | 0.1240 | .584 | .635 | -2.9 | 109 | 85.1 |  |
| **2** | **CL6** | **CL5** | 60 | 0.1531 | .431 | .492 | -2.4 | 118 | 58.3 |  |
| **1** | **CL3** | **CL2** | 158 | 0.4307 | .000 | .000 | 0.00 | . | 118 |  |





Hay que hacer esta parte…

**proc** **tree** data=pecesA n=**5** out=ClasifA;

id Obs;

**run**;

**proc** **print** data=CiudadesClasif;

**run**;