**Assignment 7**

**1 Latent Profile Analysis**

**Table 1**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Profiles | LL | FP | AIC | BIC | C-AIC |  |  | Entropy |
| 2 | -3861.875 | 16 | 7755.749 | 7823.183 | 7839.183 |  |  | .999983 |
| 3 | -3257.248 | 22 | 6558.496 | 6651.217 | 6673.217 |  |  | .999960 |
| 4 | -3254.505 | 28 | 6565.01 | 6683.019 | 6711.019 |  |  | .998873 |
| 5 | -3246.13 | 34 | 6560.259 | 6703.556 | 6737.556 |  |  | .984619 |
| 6 | -3227.739 | 40 | 6535.478 | 6704.062 | 6744.062 |  |  | .918763 |
| 7 | -3221.135 | 46 | 6534.27 | 6728.142 | 6774.142 |  |  | .886031 |
| 8 | -3218.947 | 52 | 6541.895 | 6761.054 | 6813.054 |  |  | .889544 |
| 9 | -3213.432 | 58 | 6542.863 | 6787.311 | 6845.311 |  |  | .808074 |
| 10 | -3210.392 | 64 | 6548.784 | 6818.519 | 6882.519 |  |  | .811309 |



**Figure 1**

Based on above results (Table 1 and Figure 1), 3-profiles model would the best fitting model. The 3-profiles model is characterized by the lowest BIC and C-AIC. Meanwhile, its entropy is higher than other models with more profiles.

**Table 2 Latent class marginal probabilities**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Margin | Std. Err. | [95% Conf. Interval] | |
| 1 | .22 | .0185445 | .1858138 | .25849 |
| 2 | .32 | .0208825 | .2805191 | .3622296 |
| 3 | .46 | .0223113 | .4166874 | .5039253 |
|  |  |  |  | N = 500 |

According to table 2, 22% samples can be classified to profile 1, 32% to profile 2 and 46% to profile 3.

**Table 3 Latent class marginal means**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Variables | Margin | Std. Err. | z | P>z | [95% Conf. | Interval] |
| 1 | auto | 2.563646 | .0756195 | 33.90 | 0.000 | 2.415435 | 2.711858 |
|  | compagg | 6.636324 | .0493931 | 134.36 | 0.000 | 6.539515 | 6.733133 |
|  | innov | 3.181815 | .0752397 | 42.29 | 0.000 | 3.034348 | 3.329282 |
|  | proact | 6.481761 | .0600118 | 108.01 | 0.000 | 6.36414 | 6.599382 |
|  | riskt | 6.009067 | .0710706 | 84.55 | 0.000 | 5.869771 | 6.148363 |
| 2 | auto | 3.350005 | .06331 | 52.91 | 0.000 | 3.22592 | 3.47409 |
|  | compagg | 4.062488 | .0571274 | 71.11 | 0.000 | 3.95052 | 4.174456 |
|  | innov | 2.187487 | .0626816 | 34.90 | 0.000 | 2.064634 | 2.310341 |
|  | proact | 2.393727 | .0594183 | 40.29 | 0.000 | 2.27727 | 2.510185 |
|  | riskt | 5.275005 | .0579855 | 90.97 | 0.000 | 5.161356 | 5.388655 |
| 3 | auto | 5.686956 | .0459482 | 123.77 | 0.000 | 5.5969 | 5.777013 |
|  | compagg | 1.917391 | .0479591 | 39.98 | 0.000 | 1.823393 | 2.011389 |
|  | innov | 6.447826 | .0396201 | 162.74 | 0.000 | 6.370172 | 6.52548 |
|  | proact | 4.065217 | .0507588 | 80.09 | 0.000 | 3.965732 | 4.164703 |
|  | riskt | 2.46087 | .0471058 | 52.24 | 0.000 | 2.368544 | 2.553195 |
|  |  |  |  |  |  |  | N = 500 |



**Figure 2**

Table 3 and Figure 2 reveal the composition of the profiles. Class 1 may be referred as *paranoid competitors* (22%), who exhibit high level of competitive aggressiveness (*M*=6.64), proactiveness (*M*=6.48) and risk taking (*M*=6.00), with, however, low level of innovativeness (*M=*3.18) and autonomy (*M=*2.56). Class 2 may be referred as *anxious risktaker* who is characterized by significant risk taking tendency (*M=*5.27) but by low level on other aspects. Class 3 may be referred as *low-key innovators*, who highly focus on innovativeness (*M=*6.45) and autonomy (*M=*5.68), but exhibit less aggressiveness (*M*=1.91) and risk taking (*M*=2.46).

**Table 4 Mean of ROA for each class**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | ROA Mean | Std. Err. | [95% Conf. | Interval] |
| 1 | .0490909 | .0029183 | .0433573 | .0548245 |
| 2 | .03375 | .0024229 | .0289896 | .0385104 |
| 3 | .1015217 | .0019309 | .0977281 | .1053154 |

Based on **Table 4**, the three classes firms significantly distinguish with each other on ROA performance. The 95% confidence interval of their means show no overlap with each other. The low-key innovators have best performance (*M* = 10%) over the others, while the performance of anxious risktakers is the worst (*M*=3%). It is reasonable to imply that anxious risktakers may tend to take unnecessary risk which offers no innovative benefits but only incur loss to their revenue. On the contrary, the low-key innovators foster innovational culture within their firms but constrain risky investment. Their disposition of high autonomy and low competitive aggressiveness also indicate that they might often have their own consideration about future development and can hardly be involved into irrational competition.

**2 Cluster Analysis**

2.1 Report the same information as in Part 1, but for clusters instead of profiles. However, make sure to include the distinctive key decisions you made and why you made them.

**Decision of standardization.** Both standardized and original data of variables are taken into hierarchical and non-hierarchical clustering analysis. The final results of these two methods show *no any difference* in clustering configuration. Therefore, I just take standardized version of results for report use.

**Decision of linkage method.** Ward’s method expects relative equal sizes of cluster observations. Since I finally choose a 3-groups clustering, this condition is likely be violated according to Table 5. The centroid method might be more suitable. However, the centroid method fails in drawing the dendrogram in Stata. Meanwhile, when I compare the final non-hierarchal clustering results with starting value from both centroid method and Ward’s method, there is *no any difference* in clustering configurations. Therefore, I just take the Ward’s method for report use.

**Decision of validation test.** First, I have done four different streams clustering analysis, with standard/non-standard data and Ward’s/centroid method, which implies a high reliability (with complete consistency). Second, the significant difference in ROA of each cluster (*F*=269.37, *p-value* = .000) indicates these clustering result also has a high criterion-related validity (Ketchen & Shook, 1996).

**Table 5 Clustering description**

|  |  |  |  |
| --- | --- | --- | --- |
| Class | Freq. | Percent | Cum. |
| **1** | 110 | 22.00 | 22.00 |
| **2** | 160 | 32.00 | 54.00 |
| **3** | 230 | 46.00 | 100.00 |
| Total | 500 | 100.00 |  |

**Table 6 Mean of variable for each class**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | Variable | Mean | Std. Err. | [95% Conf. | Interval] |
| 1 | z\_auto | -1.08863 | 0.048933 | -1.18477 | -0.99249 |
|  | z\_compagg | 1.521416 | 0.025183 | 1.471938 | 1.570895 |
|  | z\_innov | -0.5689 | 0.036276 | -0.64018 | -0.49763 |
|  | z\_proact | 1.466259 | 0.036482 | 1.394582 | 1.537937 |
|  | z\_riskt | 1.075841 | 0.041097 | 0.995097 | 1.156585 |
| 2 | z\_auto | -0.5816 | 0.040908 | -0.66197 | -0.50122 |
|  | z\_compagg | 0.213653 | 0.029087 | 0.156505 | 0.270802 |
|  | z\_innov | -1.04659 | 0.030177 | -1.10588 | -0.9873 |
|  | z\_proact | -1.01086 | 0.036079 | -1.08174 | -0.93997 |
|  | z\_riskt | 0.652849 | 0.033483 | 0.587064 | 0.718633 |
| 3 | z\_auto | 0.925238 | 0.029662 | 0.866961 | 0.983515 |
|  | z\_compagg | -0.87626 | 0.024396 | -0.92419 | -0.82833 |
|  | z\_innov | 1.000149 | 0.019057 | 0.962708 | 1.03759 |
|  | z\_proact | 0.00195 | 0.030793 | -0.05855 | 0.062449 |
|  | z\_riskt | -0.96869 | 0.027175 | -1.02208 | -0.9153 |



**Figure 3**

**Table 7 Mean of ROA for each class**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | ROA Mean | Std. Err. | [95% Conf. | Interval] |
| 1 | .0490909 | .0029183 | .0433573 | .0548245 |
| 2 | .03375 | .0024229 | .0289896 | .0385104 |
| 3 | .1015217 | .0019309 | .0977281 | .1053154 |

2.2 Produce a dendrogram and the Duda-Hart and Calinski-Harabasz stopping indices of the hierarchical clustering results. How many clusters do they suggest and what criteria did you apply to make this assessment?



**Figure 4**

**Table 8 Duda/Hart indices**

|  |  |  |
| --- | --- | --- |
| Number of clusters | Je(2)/Je(1) | pseudo T-squared |
| 1 | 0.3918 | 772.96 |
| 2 | 0.2995 | 626.80 |
| 3 | 0.8463 | 41.41 |
| 4 | 0.8302 | 32.31 |
| 5 | 0.7855 | 42.87 |
| 6 | 0.7993 | 27.12 |
| 7 | 0.7809 | 24.12 |
| 8 | 0.7120 | 34.38 |
| 9 | 0.7658 | 21.72 |
| 10 | 0.7566 | 22.20 |

**Table 9 Calinski-Harabasz indices**

|  |  |
| --- | --- |
| Number of clusters | Calinski/Harabasz pseudo-F |
| 2 | 772.96 |
| 3 | 1217.75 |
| 4 | 880.80 |
| 5 | 712.90 |
| 6 | 613.54 |
| 7 | 542.22 |
| 8 | 490.12 |
| 9 | 451.77 |
| 10 | 420.36 |
| 11 | 396.49 |
| 12 | 377.47 |
| 13 | 358.99 |
| 14 | 343.26 |
| 15 | 329.08 |

The dendrogram indicates that the additional grouping, after 3 clusters have been drawn, contributes very little to further revealing the dissimilarity of observations (Figure 4). The Calinski-Harabasz indices are consistent with this implication by showing the 3 clusters grouping have the best between-cluster dispersion against within-cluster dispersion (Table 9). The Duda/Hart indices reveal that the first two clusters have been performing well (Table 8). The elbow point might locate at the 3rd clustering. (I am not sure the weather the elbow point itself should be included when explaining the Duda/Hart indices.) However, considering all of these analytical findings together, I think 3-clusters grouping would be the best fit model.

2.3 How do the results from the cluster analyses compare to your latent profile analysis results?

These two analytical results are virtually identical.