Paper 224-26

The FASTCLUS Procedure as an Effective Way to Analyze Clinical Data

Lev Sverdlov, Ph.D., Innapharma, Inc., Park Ridge, NJ

ABSTRACT

This paper presents an example of the fast cluster analysis (SAS/STAT®, FASTCLUS), using data from a pilot clinical trial with a new antidepressant drug. The list of variables for cluster analysis was selected after principle component analysis (PRINCOMP procedure) from the efficacy parameters investigated in the study. Three to six clusters were modeled to get the optimal cluster structure with a proportional number of subjects per cluster and the maximum statistical difference between clusters. The best statistical solution was obtained using seven psychometric score variables and four clusters. Using canonical discriminant analysis (DISCRIM procedure), seven variables from cluster analysis were transformed into three canonical variables. The plot of any two out of three canonical variables shows a graphic display of the distinct separation between clusters.

INTRODUCTION

SAS® is the predominant statistical tool used in the pharmaceutical industry to analyze clinical data and to prepare integrated medical and statistical reports for submission to the FDA. The majority of statistical analyses of clinical trial data in the field of clinical psychiatry have been based on two-dimensional models in which the active drug group is simply compared with the placebo group. The objective of this presentation is to discuss some methodological aspects of a multivariate statistical method and demonstrate an example of how fast cluster analysis (SAS/STAT®, FASTCLUS [1]) separates a heterogeneous study population into homogeneous subgroups (clusters) with a relatively similar response to treatment within clusters and a significantly different response between clusters.

METHOD

Fifty-two subjects (26 placebo and 26 treated with a new antidepressant drug) were enrolled into a randomized, double-blind, placebo-controlled, parallel-design, single-center, pilot study [2]. Efficacy was evaluated by the Hamilton Depression Rating Scale (HAMD), Montgomery-Asberg Depression Rating Scale (MADR), Carroll Self-Rating Scale (CSRS), Clinical Global Improvement (CGI) with the last observation carried forward approach for missing data. In the SAS data set the subjects were sorted in ascending order by subject number according to the randomization table. Because of the limited sample size, all of the cluster analysis subjects were included in a training group (there was no replication group of subjects from this study but a replication group can be created in the next study). The list of variables for the cluster analysis [2,3] was selected from the four efficacy parameters investigated in the study for two treatment days (eight variables). After the principal component analysis (PRINCOMP procedure) only seven out of eight variables were selected with correlation less than 0.9. Using seven cluster variables, three to six clusters were modeled to get the optimal cluster structure. The criteria for selection of the clusters included a proportional number of subjects per cluster and the maximum statistical difference (ie. minimum chi-square probability) between clusters over the following treatment groups: placebo, drug group with plasma parent drug concentration at 1 hour after dosing equal or above the minimum projected therapeutic concentration (MPTC) and below MPTC. The corresponding statistical solutions obtained for the different number of clusters are presented in Table 1. According to Table 1, only three or four clusters can be considered a reasonable solution because for five or six clusters, one cluster represents only the effect of outliers. But the solution for four clusters has a better statistical effect of separation in terms of P-value. That's why the four cluster solution was selected as optimal.

Canonical discriminant analysis (DISCRIM procedure) transformed seven cluster variables into three canonical variables in order to plot any two canonical variables out of three to visualize the effect of separation. MACROs 1 and 2 with the FASTCLUS procedure for training and the replication group of subjects are presented in the Appendix. In addition MACRO 3 can be

used for the replication variables, either for training or the replication group of subjects.

RESULTS

The input data set (Table 2) contained subject number, type of treatment, plasma group (drug concentration in plasma ≥ MPTC or < MPTC) and psychometric assessments for the cluster variables. For the description of all variables in Table 2, see Appendix. Table 3 shows the output from FASTCLUS procedure with all parameters (Replace=full, Radius=0, Maxclusters=4, Maxiter=20). Table 4 presents cluster definition and distance for all subjects.

DISCUSSION

Figure 1 presents a plot of two of the three canonical variables for each of the four clusters. The plot shows a display of the distinct separation between clusters. The analysis of cluster structure (see Table 3) shows that clusters 2 and 3 collected responders for drug and clusters 1 and 4 collected non-responders for drug with different magnitude of response by cluster. After separation with FASTCLUS the variability of response among drug treated subjects by cluster was attributed to a pharmacokinetic effect (concentration of drug in plasma). Placebo subjects contributed equally (25%) to all four clusters. The distribution of baseline characteristic was similar, and there was no difference between clusters in all demographic features. Cluster analysis to replicate platelet serotonin uptake rates as a biochemical marker of treatment effect was reported in [4].

SUMMARY

SAS/STAT® is a powerful and flexible tool for the multivariate statistical analysis of clinical data. The FASTCLUS procedure effectively separated a heterogeneous study population of patients diagnosed with major depression into relatively homogeneous subgroups. The division of the study population into four clusters, two of which were classified as clusters with responders to treatment and two that were considered as clusters with non-responders to treatment, as well as the equal distribution of placebo subjects in all four clusters, provides substantial evidence of the value of cluster analysis with the FASTCLUS procedure. All the methodological aspects in this presentation and MACROs for the training and replication group of subjects can be used as an effective approach in drug development for many clinical trials in the CNS therapeutic area.

REFERENCES

1. SAS/STAT®, User's Guide, Version 6, Fourth Edition, 53-101, 823-850.
2. Feighner, J.P., Ehrensing, R.H., Kastin, A.J., Noble, J.F., Sverdlov, L., Abajian, H., Nicolau, G. (2000), "A double-blind, placebo-controlled, efficacy, safety, and pharmacokinetic study of INN 00835, a novel antidepressant peptide, in the treatment of major depression", *Journal of Affective Disorders*, *61.*
3. Ravindra, Khattree, and Dayanand, N. Naik, (2000), "Multivariate data reduction and discrimination with SAS® Software", *Cary, NC: SAS Institute Inc.*
4. Sverdlov, L, Noble, J.F., Nicolau, G., (1999), "Cluster analysis (FASTCLUS Procedure) to replicate platelet serotonin uptake rates as a biochemical marker of treatment effect for a new antidepressant drug", *Proceedings of PharmaSUG '99. New Orleans, Louisiana, 161-164.*

TRADEMARKS

CONTACT INFORMATION

Lev Sverdlov, Ph.D., Innapharma, Inc.
1 Maynard Drive, Suite 205,
Park Ridge, NJ 07656
Phone: (201) 505-1300 Ext. 645
E-MAIL: LSVERDLOV@AOL.COM

TABLE 1

No. of Clusters	Chi-Square P-Value of Separation	Minimum No. of Subjects Per Cluster	Maximum No. of Subjects Per Cluster
3	0.063	13	19
4	0.049	10	15
5	0.092	2	13
6	0.174	1	13

TABLE 2

OBS	SUBJ	TREAT	GRP	PHAMD_7	PHAMD_14	MADS_14	PCSRS_7	PCSRS_14	PCGI_7	PCGI_14
1	001	Placebo	placebo	-12.500	-31.250	-37.838	-43.750	-81.250	0.0000	-20.0000
2	002	Placebo	placebo	-61.765	-38.235	-47.368	-34.483	-6.897	-40.0000	-40.0000
3	003	INN00835	con<5	-35.484	-32.258	-38.889	-45.455	-51.515	-20.0000	-20.0000
4	004	INN00835	con<5	-44.000	-28.000	-40.741	-78.261	-43.478	-25.0000	-25.0000
5	006	Placebo	placebo	-91.667	-94.444	-92.857	-89.286	-100.000	-83.3333	-83.3333
6	007	Placebo	placebo	-14.815	3.704	11.538	0.000	18.182	0.0000	0.0000
7	008	INN00835	con<5	0.000	-12.500	-11.765	-4.348	0.000	0.0000	0.0000
8	009	Placebo	placebo	-69.697	-78.788	-69.444	-56.667	-73.333	-40.0000	-40.0000
9	010	INN00835	con>=5	-55.556	-59.259	-50.000	-61.905	-52.381	-25.0000	-25.0000
10	011	Placebo	placebo	-28.571	7.143	3.125	-32.000	0.000	-20.0000	0.0000
11	012	Placebo	placebo	-23.810	9.524	-3.125	-90.000	-55.000	-25.0000	0.0000
12	013	Placebo	placebo	-61.290	-58.065	-66.667	-67.857	-67.857	-60.0000	-60.0000
13	014	INN00835	con<5	-77.778	-86.111	-91.667	-75.000	-69.444	-60.0000	-60.0000
14	015	INN00835	con<5	-26.667	13.333	5.714	-46.429	-7.143	-20.0000	0.0000
15	016	INN00835	con<5	-10.345	3.448	-5.128	-7.692	-7.692	0.0000	0.0000
16	017	Placebo	placebo	-47.059	-35.294	-51.351	-25.000	-29.167	-40.0000	-40.0000
17	018	Placebo	placebo	-20.000	-20.000	-20.000	-41.935	-58.065	-20.0000	-20.0000
18	019	INN00835	con<5	-14.815	-22.222	-6.667	-30.435	-4.348	0.0000	0.0000
19	020	INN00835	con<5	-40.741	-22.222	-28.125	-25.000	5.000	-40.0000	-20.0000
20	021	Placebo	placebo	-44.444	-29.630	-25.000	-65.385	-23.077	-25.0000	0.0000
21	022	Placebo	placebo	-75.862	-82.759	-94.444	-75.000	-84.375	-80.0000	-80.0000
22	023	INN00835	con>=5	-90.909	-81.818	-96.552	-94.444	-94.444	-75.0000	-75.0000
23	024	Placebo	placebo	-59.375	-59.375	-62.162	-77.778	-81.481	-40.0000	-40.0000
24	025	INN00835	con<5	-54.839	-61.290	-60.606	-62.963	-70.370	-40.0000	-40.0000
25	026	Placebo	placebo	-67.857	-67.857	-69.444	-4.000	-44.000	-25.0000	-25.0000
26	027	Placebo	placebo	-86.667	-80.000	-85.714	-100.000	-88.462	-80.0000	-60.0000
27	028	INN00835	con>=5	-72.000	-64.000	-67.568	-77.419	-80.645	-60.0000	-40.0000
28	029	INN00835	con>=5	-100.000	-96.875	-100.000	-97.436	-94.872	-80.0000	-80.0000
29	031	INN00835	con<5	-50.000	-70.000	-82.353	-52.941	-58.824	-40.0000	-60.0000
30	032	Placebo	placebo	8.000	0.000	-25.000	0.000	18.182	0.0000	-20.0000
31	033	INN00835	con>=5	-52.000	-72.000	-75.000	-52.174	-73.913	-50.0000	-75.0000
32	034	INN00835	con>=5	-87.879	-100.000	-100.000	-82.609	-100.000	-60.0000	-80.0000
33	035	Placebo	placebo	-28.571	-38.095	-28.000	-50.000	-13.636	0.0000	-25.0000
34	036	INN00835	con<5	-58.824	.	.	-40.625	.	-20.000	.
35	037	INN00835	con<5	-36.000	-8.000	-21.875	-27.273	-4.545	-20.0000	-20.0000
36	039	INN00835	con<5	-40.000	-36.000	-44.828	-88.889	-72.222	-25.0000	-25.0000
37	040	Placebo	placebo	-27.273	-21.212	-29.730	-39.474	-18.421	-20.0000	-20.0000
38	041	INN00835	con<5	-86.486	-78.378	-78.378	-67.568	-75.676	-60.0000	-60.0000
39	042	Placebo	placebo	-12.000	-24.000	-19.355	0.000	5.556	-20.0000	-20.0000
40	043	INN00835	con>=5	-100.000	-95.652	-100.000	-65.385	-92.308	-75.0000	-75.0000
41	044	INN00835	con<5	13.793	6.897	9.091	16.667	6.667	0.0000	0.0000
42	045	Placebo	placebo	-46.154	-19.231	-13.333	-13.793	6.897	-40.0000	-20.0000
43	046	Placebo	placebo	-75.000	-70.833	-68.966	-79.167	-75.000	-50.0000	-50.0000
44	047	INN00835	con>=5	-43.333	-20.000	-24.324	-27.778	-30.556	-20.0000	-20.0000
45	048	Placebo	placebo	-77.273	-81.818	-80.000	-78.947	-78.947	-50.0000	-50.0000
46	049	Placebo	placebo	-45.161	-3.226	11.765	-26.316	73.684	0.0000	0.0000
47	050	INN00835	con>=5	-89.286	.	.	-73.684	.	-80.0000	.
48	051	INN00835	con>=5	-62.500	.	.	-34.783	.	-40.0000	.
49	052	Placebo	placebo	-70.370	-51.852	-51.613	-45.455	-27.273	-60.0000	-40.0000

TABLE 3

FASTCLUS Procedure: Replace=FULL Radius=0 Maxclusters=4 Maxiter=20 Converge=0.02

Initial Seeds

Cluster	PHAMD_7	PHAMD_14	PMADS_14	PCSRS_7	PCSRS_14	PCGI_7	PCGI_14
1	-23.810	9.524	-3.125	-90.000	-55.000	-25.000	0.000
2	-67.857	-67.857	-69.444	-4.000	-44.000	-25.000	-25.000
3	-100.00	96.875	-100.00	-97.436	-94.872	-80.000	-80.000
4	-45.161	-3.226	11.765	-26.316	73.684	0.000	0.000

Minimum Distance Between Initial Seeds = 142.0459

Relative Change in Cluster Seeds

Iteration	Criterion	1	2	3	4
1	25.4795	0.3843	0.3520	0.2640	0.5031
2	14.5161	0.0743	0.0766	0.0241	0.0421
3	14.2144	0	0	0	0

Convergence criterion is satisfied.

Criterion Based on Final Seeds = 14.214

Cluster Summary

Cluster	Frequency	RMS Std Deviation	Maximum Distance from Seed to Observation	Nearest Cluster	Distance Between Cluster Centroids
1	10	15.4036	58.8475	2	62.0776
2	11	15.3488	49.8689	1	62.0776
3	15	12.7086	47.2314	2	74.8394
4	13	16.3105	74.8371	1	75.6822

Statistics for Variables

Variable	Total STD	Within STD	R-Squared	RSQ/(1-RSQ)
PHAMD_7	28.965680	14.043915	0.779616	3.537533
PHAMD_14	33.827598	13.232222	0.857190	6.002288
PMADS_14	34.443129	12.849468	0.870102	6.698358
PCSRS_7	30.046986	18.053130	0.661566	1.954783
PCSRS_14	40.836301	19.918186	0.777954	3.503565
PCGIS_7	25.781988	12.509147	0.779304	3.531126
PCGIS_14	26.621338	11.280262	0.832423	4.967392
OVER-ALL	31.784679	14.856706	0.795682	3.894327

Pseudo F Statistic=58.41 Approximate Expected Over-All R-Squared=0.4161

Cubic Clustering Criterion = 25.707

WARNING: The two above values are invalid for correlated variables.

Cluster Means

Cluster	PHAMD_7	PHAMD_14	PMADS_14	PCSRS_7	PCSRS_14	PCGI_7	PCGI_14
1	-31.9415	-24.6921	-29.2474	-57.0926	-44.7220	-18.0000	-17.5000
2	-59.8037	-57.9945	-60.4825	-45.1453	-49.3028	-37.2727	-38.8889
3	-81.6064	-81.6253	-85.5580	-78.3984	-83.9960	-66.8889	-66.3095
4	-19.4981	-5.9136	-6.9242	-15.1245	8.4952	-12.3077	-7.6923

Cluster Standard Deviations

Cluster	PHAMD_7	PHAMD_14	PMADS_14	PCSRS_7	PCSRS_14	PCGI_7	PCGI_14
1	11.1994	13.6217	12.2636	22.0455	22.9714	9.7753	9.5015
2	7.6674	14.2811	11.6495	20.4869	24.6832	10.8082	10.2402
3	13.3732	12.5294	12.6062	12.9710	11.0580	12.4860	13.7620
4	19.6954	12.9470	14.2285	17.6530	21.3111	15.3590	10.1274

TABLE 4

OBS	SUBJID	TREAT	GRP	CLUSTER	DISTANCE
1	001	Placebo	Placebo	1	48.3461
2	003	INN00835	con<5	1	18.8306
3	004	INN00835	con<5	1	29.0403
4	012	Placebo	Placebo	1	58.8475
5	018	Placebo	Placebo	1	25.8484
6	021	Placebo	Placebo	1	33.0342
7	035	Placebo	Placebo	1	39.8676
8	039	INN00835	con<5	1	48.0423
9	040	Placebo	Placebo	1	32.3505
10	047	INN00835	con>=5	1	35.3033
11	002	Placebo	Placebo	2	49.8689
12	009	Placebo	Placebo	2	36.4614
13	010	INN00835	con>=5	2	27.6298
14	017	Placebo	Placebo	2	39.7629
15	024	Placebo	Placebo	2	45.9775
16	025	INN00835	con<5	2	28.3815
17	026	Placebo	Placebo	2	48.0315
18	031	INN00835	con<5	2	36.3744
19	036	INN00835	con<5	2	27.3141
20	051	INN00835	con>=5	2	16.8785
21	052	Placebo	Placebo	2	35.0891
22	006	Placebo	Placebo	3	35.4078
23	013	Placebo	Placebo	3	42.2325
24	014	INN00835	con<5	3	19.5617
25	022	Placebo	Placebo	3	22.0062
26	023	INN00835	con>=5	3	26.7470
27	027	Placebo	Placebo	3	26.9547
28	028	INN00835	con>=5	3	38.4503
29	029	INN00835	con>=5	3	40.2429
30	033	INN00835	con>=5	3	47.2314
31	034	INN00835	con>=5	3	33.0800
32	041	INN00835	con<5	3	18.9659
33	043	INN00835	con>=5	3	33.5172
34	046	Placebo	Placebo	3	32.6829
35	048	Placebo	Placebo	3	25.0347
36	050	INN00835	con>=5	3	24.3014
37	007	Placebo	Placebo	4	31.4408
38	008	INN00835	con<5	4	29.0851
39	011	Placebo	Placebo	4	28.7937
40	015	INN00835	con<5	4	43.8680
41	016	INN00835	con<5	4	26.5062
42	019	INN00835	con<5	4	29.9662
43	020	INN00835	con<5	4	46.8485
44	032	Placebo	Placebo	4	41.7539
45	037	INN00835	con<5	4	32.0700
46	042	Placebo	Placebo	4	31.3991
47	044	INN00835	con<5	4	52.4744
48	045	Placebo	Placebo	4	43.0305
49	049	Placebo	Placebo	4	74.8371

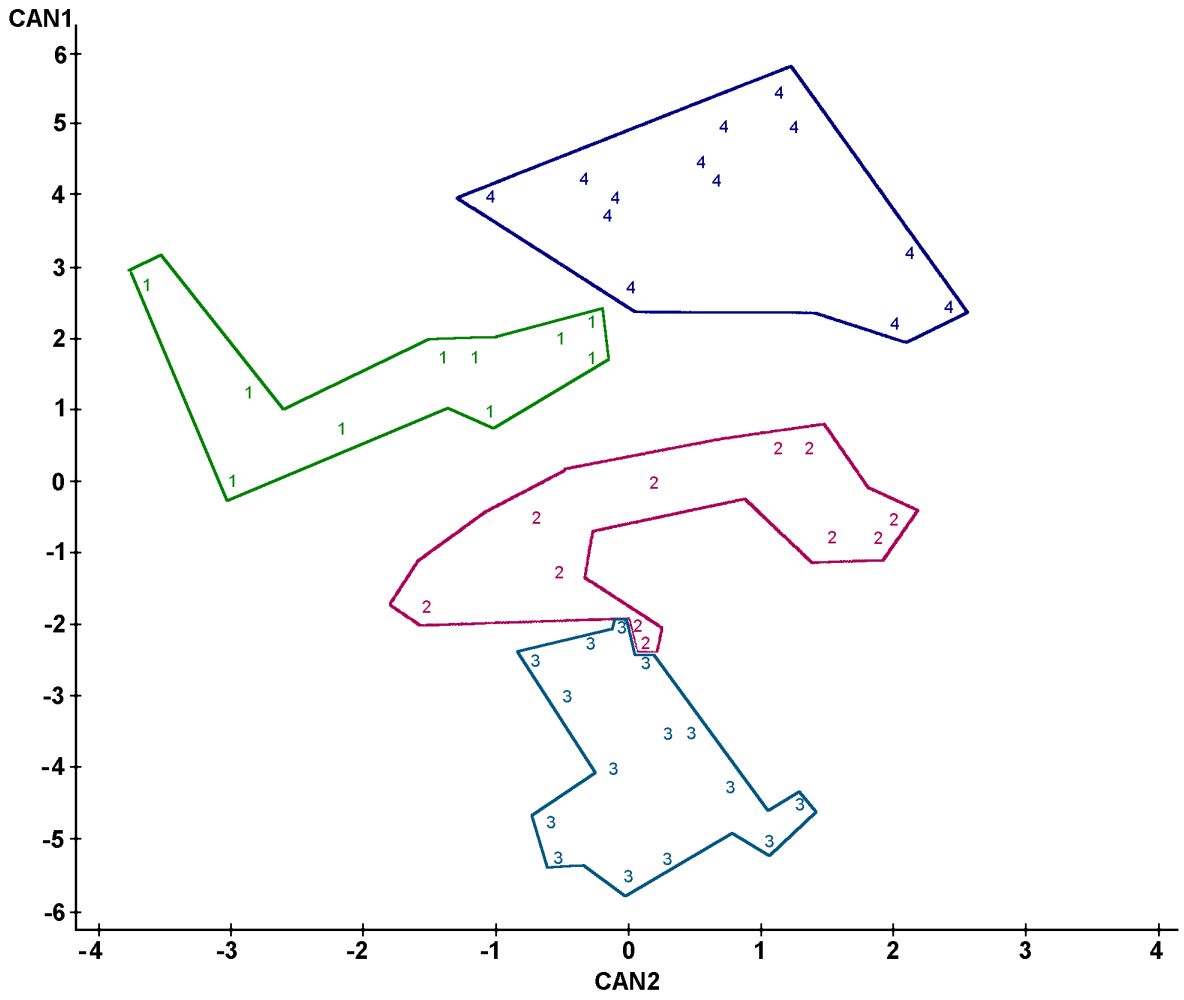

FIGURE 1

Numbers 1, 2, 3 and 4 mean corresponding cluster membership for each of the subjects.

APPENDIX 1

```
/*****************************************/
/**  MACRO 1 for Cluster Analysis **/
/**     Training Group of Subjects   **/
/*****************************************/

  title 'Training Group of Subjects' ;

  %macro clus_tr(fl_tr,  max_cl) ;

  proc fastclus data=&fl_tr maxc=&max_cl
                  out=clus maxiter=20  ;
  var phamd_7 phamd_14 pmadr_14
       pcsrs_7 pcsrs_14 pcgis_7 pcgis_14 ;
  run ;

  proc sort data=clus ;
  by cluster ;
  run ;

/*  Print all Subjects with cluster definition  */
  proc print data=clus ;
  var subjid subjinit grp cluster distance ;
  run ;

  proc freq data=clus ;
  table cluster * grp / chisq exact ;
  run ;

/* Output with cluster structure for Replication group  */
  proc means data=clus ;
  by cluster ;
  var phamd_7 phamd_14 pmadr_14
      pcsrs_7 pcsrs_14 pcgis_7 pcgis_14 ;
  output out=replic mean=phamd_7 phamd_14
                pmadr_14 pcsrs_7 pcsrs_14 pcgis_7 pcgis_14 ;
  run ;

  /* Canonical Discriminant Analysis */
  proc candisc data=clus anova out=can  ;
  class cluster ;
  var phamd_7 phamd_14 pmadr_14
      pcsrs_7 pcsrs_14 pcgis_7 pcgis_14
  run ;

  proc print data=can ;
  var subjid subjinit grp cluster can1 can2 can3 ;
  run ;

  /*  Plots of two of the three canonical variables  */
  proc plot data=can ;
  plot can1*can2=cluster / haxis=-4 to 4 by 1
                           vaxis=-6 to 6 by 1 ;
  run ;

  proc plot data=can ;
  plot can1*can3=cluster / haxis=-4 to 4 by 1
                           vaxis=-6 to 6 by 1 ;
  run ;

  proc plot data=can ;
  plot can2*can3=cluster / haxis=-4 to 4 by 1
                           vaxis=-6 to 6 by 1 ;
  run  ;

  %mend clus_tr ;
```

```
/*****************************************/
/**  MACRO 2 for Cluster Analysis **/
/**  Replication Group of Subjects **/
/*****************************************/

  title 'Replication Group of Subjects'  ;

  %macro clus_rp(fl_rp, max_cl)  ;

  proc fastclus data=&fl_rp
  seed=replic replace=none maxiter=0 maxc=&max_cl
          out=clus_rp  ;
  var phamd_7 phamd_14 pmadr_14
       pcsrs_7 pcsrs_14 pcgis_7 pcgis_14 ;
  run ;

  proc sort data=clus_rp ;
  by cluster ;
  run ;

  proc freq data=clus_rp ;
  table cluster * grp / chisq exact ;
  run ;

 %mend clus_rp ;

/*****************************************/
/**  MACRO 3 for Cluster Analysis **/
/**     Training or Replication       **/
/**       Group of Subjects,          **/
/**      Replication Variables        **/
/*****************************************/

  title 'Replication Variables'              ;

%macro var_rp(&fl, &rp )  ;

  proc means data=&fl n nmiss mean std stderr
                min max range ;
  var &rp ;
  by cluster ;
  run ;

 %mend var_rp ;
```

Definition of Terms in Appendix

fl_tr	- SAS input data set for training group of subjects.
fl_rp	- SAS input data set for replication group of subjects.
max_cl	- Number of clusters.
clus_tr	- SAS output data set for training group of subjects.
clus_rp	- SAS output data set for replication group of subjects.
grp	- Treatment groups.
phamd_7	- Percent change from baseline for HAMD (Hamilton 21 item Depression Rating Scale) for Day 7.
phamd_14	- Percent change from baseline for HAMD for Day 14 (Week 1 for follow-up period).
pmads_14	- Percent change from baseline for MADS (Montgomery-Asberg Depression Scale) for Day 14.
pcsc_7	- Percent change from baseline for CSRS (Carroll Self-Rating Scale for Depression) for Day 7.
pcsc_14	- Percent change from baseline for CSRS for Day 14.
pcgis_7	- Percent change from baseline for CGI (Clinical Global Impression - Severity of Illness) for Day 7.
pcgis_14	- Percent change from baseline for CGI for Day 14.