

Cluster Analysis Explained – Lincoln Sheep Example

In June, 2007, the National Lincoln Sheep Breeders Association (NLSBA) provided NAGP with access to their database. There were 13,458 records from animals born from 1976-2007. Of these, 8,972 (66.7%) were ewes and 4,486 (33.3%) were rams. Birth years were unknown for 12.9% of the data. The only data edits needed were for 2 rams that appeared as their own sire. Their sire was converted to unknown. Sires were known for 89.0% of the animals and dams were known for 88.9%. There have been 1,136 sires and 4,501 dams during the lifetime of the database.

The cluster analysis included the relationship between the 149 rams that sired the 2006-2007 lamb crop. These are the males assumed to be available from which to collect. If they are not available, it is known they have offspring, so a son might be available as a replacement. The Cluster procedure in SAS (Version 9.2; SAS, 2009) was used to cluster the rams using the Ward method. T-statistics were examined for natural breaks in the groupings to determine the appropriate number of clusters. The analysis resulted in 16 distinct groups of rams. The average relationship between the 149 rams was 2.4%. Table 1 shows the average relationship within each of the 16 clusters.

Table 1. Within Cluster Relationship of 149 Lincoln Rams

Cluster	n	Avg. Rel.
1	12	0.12
2	5	0.40
3	12	0.15
4	4	0.49
5	8	0.14
6	8	0.21
7	20	0.12
8	5	0.32
9	12	0.12
10	3	0.42
11	5	0.38
12	11	0.17
13	3	0.42
14	11	0.25
15	7	0.29
16	23	0.01

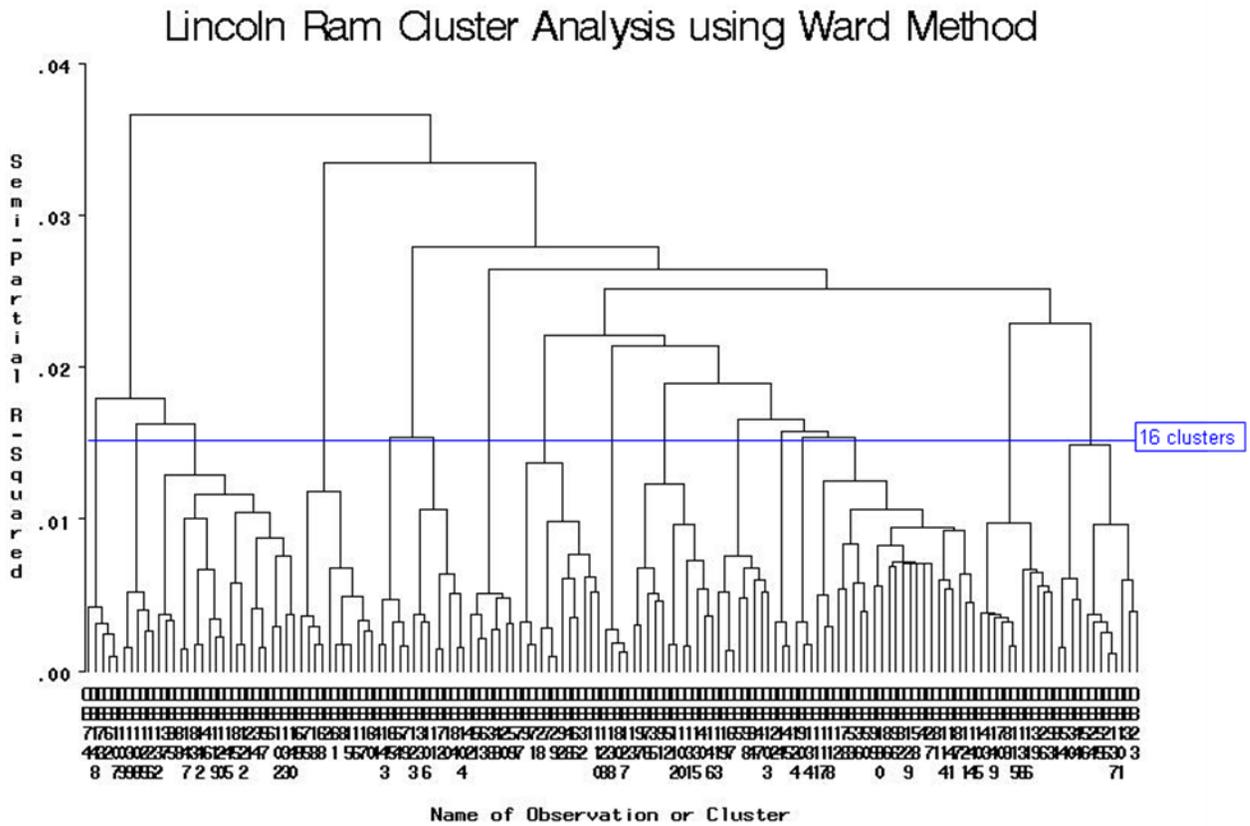
All of the clusters have a high average relationship with the exception of cluster 16. To put the relationships into perspective, a 0.25 relationship is the equivalent of a half-sib and 0.125 is equivalent to a cousin. Cluster 16 is a cluster where the animals that don't fit anywhere else got placed. They would be expected to have a high level of genetic diversity and should be sampled heavier than other clusters.

The clusters not only have a high degree of relatedness within each cluster, but a low degree of relatedness between clusters. This can be seen in Table 2. A graphical representation of the clusters is shown in Figure 1.

Table 2. Between Cluster Relationship of 149 Lincoln Rams

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.12	0.02	0.00	0.00	0.00	0.02	0.04	0.03	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.00
2		0.40	0.00	0.00	0.00	0.01	0.10	0.05	0.01	0.00	0.02	0.02	0.01	0.00	0.01	0.00
3			0.15	0.00	0.01	0.02	0.01	0.01	0.01	0.00	0.03	0.03	0.03	0.01	0.00	0.00
4				0.49	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01
5					0.14	0.03	0.02	0.02	0.03	0.00	0.00	0.02	0.01	0.01	0.05	0.01
6						0.21	0.04	0.04	0.02	0.00	0.11	0.01	0.02	0.05	0.03	0.01
7							0.12	0.08	0.02	0.00	0.04	0.03	0.02	0.00	0.01	0.01
8								0.32	0.01	0.00	0.04	0.03	0.02	0.00	0.01	0.02
9									0.12	0.00	0.02	0.04	0.01	0.00	0.01	0.01
10										0.42	0.00	0.00	0.00	0.00	0.00	0.00
11											0.38	0.01	0.01	0.02	0.00	0.01
12												0.17	0.02	0.00	0.01	0.01
13													0.42	0.03	0.01	0.00
14														0.25	.01	0.00
15															0.29	0.01
16																0.01

Figure 1. Lincoln Cluster Graph Showing 16 Clusters



Literature Cited:

SAS Institute Inc. 2009. Base SAS® 9.2 Procedures Guide. Cary, NC: SAS Institute Inc.