

Accounting for Management in Cost of Livestock Disease Studies

Stephen L. Ott
Centers for Epidemiology and Animal Health
USDA:APHIS:VS
Natural Resources Research Center, Bldg. B, 2E5
2150 Centre Ave.
Fort Collins, CO 80526-8117

Email: Stephen.L.Ott@aphis.usda.gov

Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, CO, August 1-4, 2004.

Any opinions and conclusions are those of the author and do not necessarily reflect the views of the U.S. Department of Agriculture.

Copyright 2004 by U.S. Department of Agriculture. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract:

Assessing the impacts of livestock disease on productivity is enhanced when the management abilities of producers are taken into consideration. One method of doing this is to develop a management index of several production/marketing practices using correspondence analysis which can combine several practices into a few variables. Also, through correspondence analysis herds can be divided into better managed and more poorly managed herds to test the association of management on herd health, productivity, and profitability.

Key Words: Correspondence analysis, management, herd health, herd productivity, herd profitability.

Accounting for Management in Cost of Livestock Disease Studies

Introduction

When analyzing the impact of a disease on livestock productivity in a cross-sectional study, the analyst needs to account for the innate ability of the farm manager to organize production and minimize the impact of disease. If the analyst finds a negative relationship between disease prevalence and livestock productivity, the question becomes, “Is low productivity truly associated with the disease or is disease prevalence merely a proxy measure of the manager’s ability to produce livestock products?” In other words, good managers, because of their innate ability to manage livestock, will have low prevalence of disease and high livestock productivity. If one accepts this latter statement as true, then it becomes imperative to quantify the innate management ability of livestock producers when attempting to measure the impact of disease on livestock productivity.

Unfortunately, measuring a person’s innate management ability is impossible to do. Instead, what are often measured are the characteristics of the manager: age, education level, years of experience, etc. Higher education levels and more years of experience and age are assumed to be associated with better managers. Another way to measure a manager’s ability is to record the number of “good management practices” that a manager uses. Examples of some good management practices include the use of farm records, feeding practices, and vaccination practices. Several management practice variables could be used in the analysis. Since several good management practices can be employed simultaneously, one would like to include them all

in any analysis, but to do so may result in multicollinearity problems. What is needed is some type of variable reduction method so that all of the management practices can be incorporated into the analysis and not just a select few.

Dohoo, et al. provided an overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. Techniques included combining variables through the use of indices, principal components analysis, and factor analysis. A limitation of developing an index is the weighting of each individual management practice. Should use of computer farm records have the same weight as vaccinating for *Salmonella*? Principal components analysis and factor analysis do calculate relative loadings for each variable. Unfortunately, both principal components and factor analysis are not well suited for categorical data, especially of the 0,1 type associated with use of specific management practices. A multivariate method for analysis qualitative (categorical) data is correspondence analysis (Greenacre, 1984).

Correspondence Analysis

General Theory

“Correspondence analysis is a technique for displaying the rows and columns of a data matrix (primarily, a two-way contingency table) as points in dual low-dimensional vector spaces” (Greenacre, 1984). Basic to the understanding of correspondence analysis is the concept of a profile (Greenacre, 1993). In a contingency table, a profile is simply the individual frequencies divided by the total frequency. Each row and column in a contingency table has its own profile which are called row and column profiles, respectfully. A row profile describes the row in terms of frequencies in each of the columns and vis-a-vis for column profiles. Describing a row

(column) in terms of frequencies with each column (row) yields a joint graphical display that produces two dual displays whose row and column geometries have similar interpretations which facilitates analysis and detection of relationships (Hoffmand and Franke, 1986).

For further discussion we will assume that we are interested in describing the differences among the rows. To simplify things we will assume only three columns of attributes for six rows (6x3 contingency table). As already stated, each row's profile can be described by its frequencies in each column. If each column were an axis we could then plot the profile of each of the rows in three dimensional space. To assist us in our plotting of points we assume each axis representing a column is a unit vector so instead of positioning by frequencies, row profiles are position by percentages. All the profiles (points) would lie on a two-dimensional plane defined by the triangle that joins the extreme points of each of the axes (columns). These extreme points are called vertices. The profile of each row can then be described by its distance from each of the vertices which equals its weighted average of the vertices (Greenacre, 1993).

Profiles can be described by their distance from the weighted center of the plane which is called the centroid or average profile. The centroid is the profile point of the row of column totals. The centroid also equals the weighted sum of the row profiles, where the weights equal each row's relative frequency. Row weights are also known as the mass of each row. The greater the distance the row profiles are from the centroid, the more inertia the data set is said to have; and the greater the inertia the more differences there are among the rows (Greenacre, 1993). It should be noted that distances are measured as Chi-square distances and not as straight line or Euclidean distances. Finally, what has been said about rows is also true of columns.

We now expand to n -columns which results in row profiles lying in $n-1$ dimensional space. Since it is difficult to observe or even imagine such points in more than three-dimensional space, it becomes necessary to reduce the number of dimensions to a maximum of three. Reducing dimensions comes at cost of loss of information. The goal, then, is to reduce the dimension of the problem while keeping the loss of information to a minimum.

The profiles and vertices are in “cloud” in $n-1$ dimensional space. The idea is to find a lower dimension that captures most of the information represented by this cloud. Assume one-dimension space (a line). The goal is to find a line that comes closest to all the points, both profiles and vertices. Closeness is defined as the perpendicular Chi-square distance from the points to the line. Mathematically it is easier to minimize the sum of squared distances which is a least-squares problem. In addition we need to account for the relative importance of each profile and vertex which is accomplished by including the mass associated with each profile and vertex. Thus, the criterion used in correspondence analysis to determine lower profiles is a weighted-sum-of-squared-distances criterion. The quality of this dimension reduction can be measured by the percentage of total inertia captured by the lower-dimension projection. The greater the percentage of inertia captured, the better the quality of the dimension reduction (Greenacre, 1993). For example, if one had 11 variables there would be 10 dimensions. The first dimension might capture 30 percent of the inertia which would be three times the amount if the inertia were evenly distributed among the 10 dimensions. The second dimension might capture 20 percent of the inertia, allowing one to capture 50 percent of the inertia with just two dimensions with low probability that the two dimensions were highly correlated.

Typically, analysts focus on the first two dimensions as they form a plane which can be easily graphed.

As a Tool In Variable Reduction

Besides plotting the first two dimensions' axes, correspondence analysis can reduce the number of variables in regression models. As already stated several variables of management practices could be highly correlated with one another and cause multicollinearity problems in regression models. Having one or two dimension variables from correspondence analysis that captures most of the inertia of these management variables can be most useful in epidemiologic studies.

Since the goal is to have a dimension coefficient for each observation we no longer use a contingency table, but instead use information from each observation in the data set. Each observation is a row and the management practices of interest are the columns. But first we need to "double" the columns as the columns contain only the positive or "yes" response. The negative or "no" response needs to be recorded as well in order to take into account that the variables are bipolar. Doubling then creates symmetry between the two "poles" of each binary variable. When this is done correspondence analysis is then invariant with the direction in which we choose to scale the data (Greenacre, 1984 and Hoffman and Franke). Double coding can be best demonstrated with an example using a variable that measures sex. The variable SEX may be coded 1=male and 0=female. With doubling there are two variables for recording sex, the variables MALE and FEMALE. MALE=1 when SEX=1 and MALE=0 when SEX=0 and

FEMALE=0 when SEX=1 and FEMALE=1 when SEX=0. The doubling of variables does not change the maximum number of dimension which is still $n-1$ and not $2n-1$.

A correspondence analysis algorithm, such as PROC CORRESP (SAS Institute) generates $n-1$ dimension coordinate values for each $2n$ management variables (i.e. a $2n, n-1$ matrix). Each observation has a $2n$ vector of 1's and 0's for representing the presence/nonpresence of each management variables. The matrix and vector are multiplied which results in $n-1$ dimension scores for each observation with values generally ranging from ± 1 . These dimension scores can then be merged with the original data set. This allows a traditional regression model to be developed that can include up to $n-1$ dimension variables as explanatory or independent variables in the regression equation. (See Appendix 1 for example SAS code.)

An Example of Variable Reduction in Dairy Farm Management

U.S. Department of Agriculture's National Animal Health Monitoring Systems (NAHMS) conducts surveys of livestock producers. In 1996, NAHMS surveyed 2,542 dairy producers in the top 20 dairy states (USDA, 1996a). Phase II of the the Dairy '96 survey was completed by 1,219 of the 2,542 producers (USDA, 1996b). An objective of Dairy '96 was to determine productivity impacts of selected dairy livestock diseases. Productivity impacts associated with Johnes disease (Ott, Wells, and Wagner) and subclinical mastitis (Ott and Novak) have been investigated.

In both analyses the Dairy Herd Improvement Association (DHIA) records were used as a proxy variable for general dairy management ability. Many other management practices were collected in Dairy '96 and 23 additional management practices were considered for this demonstration of correspondence analysis. DHIA and the additional 23 management practices were included in the survey as they were recognized as being part of “best management practices” for a dairy farm. The additional practices survey were use of on-farm computer records; use of total mixed rations (tmr); testing of forages; estrus synchronization; participation in milk quality assurance program; don't allow nursing of calves; feed 4 quarts or more of colostrum; determine culling by breakeven analysis; vaccinations for bovine viral diarrhea, infectious bovine rhinotracheitis, parainfluenza virus, bovine respiratory syncytial virus, *Hemophilus somnus*, leptospirosis, *Salmonella*, *E. coli* mastitis, clostridia, and brucellosis; and finally, other preventive practices of using selenium, ionophores, vitamins A-D-E, dewormers, and probiotics. A total of 1,076 producers provided information for all 24 management practices.

The subclinical mastitis study will be used to demonstrate the impact of incorporating several best management practices through correspondence analysis. Three levels of bulk tank somatic cell count (BTSCC) (low, <200k/ml; median, 200k-399k/ml; and high, 400k+/ml) were used as measures of subclinical mastitis (Ott and Novak). Other explanatory variables taken from Ott and Novak include herd size, location (4 regions), intensive grazing of summer pasture, bovine somatotropin (BST) use, if average number of days cows were dry was in excess of 70 days, and percent Holstein breed. SUDAANTM PROC REGRESS algorithm (Shah, et al.) was used in order to incorporate sample study design into the analysis. This base model explained 41 percent of the variance associated with milk production per cow (table 1). The addition of DHIA data

use increased the model R^2 to 47 percent. Combining all 24 best management variables into a simple index that sums the number of practices used per herd also resulted in a model R^2 of 47 percent. SAS PROC CORRESP (SAS Institute) was used to calculate dimension scores. The first and second dimension scores for each observation were then inputted into the SUDAANTM regression model. The first dimension from correspondence analysis captured 20 percent of the inertia associated with the 24 management variables and focused on any use of the management practices. The second dimension captured another 8 percent and focused on vaccination use. When the first dimension was used as a proxy for management model R^2 was 45 percent, but when the second dimension was added to the model R^2 rose to 50 percent. Including the first two dimensions increased the amount of explained variance by 15 percent over the model with no management and 5 percent over the model with DHIA serving as the proxy for good management.

Another way to demonstrate the impact of accounting for management practices is to compare the coefficients and standard errors of the model's explanatory variables. In the base model BTSCC-median herds produced 993 pounds of milk per cow less than BTSCC-low herds (table 2). For BTSCC-high herds the difference from BTSCC-low herds was 2577 pounds of milk per cow. Adding management as an explanatory variable reduced these impacts. When the DHIA variable was added to the model the differences in milk production were 853 pounds for BTSCC-median and 2194 pounds for BTSCC-high. The impact on BTSCC variables was even greater when management was measured by the first two dimensions from correspondence analysis. When the first two management dimensions were added to the model BTSCC-median herds produced 782 pounds of milk per cow less than BTSCC-low herds and for BTSCC-high

herds the decline in milk production was 1977 pounds per cow. The addition of the first two management dimensions reduced the impact associated with BTSCC-median herds by 21 percent and for BTSCC-high herds by 23 percent.

The impact of adding management variables was equally impressive on some of the other explanatory variables. The coefficient of milk production for herd size (log of number of cows) was 1002 pounds and declined to 796 pounds, when DHIA variable was added and only 484 pounds, when management was measured by the first two dimensions from correspondence analysis (table 2). Without management variables the coefficient of milk production for the western region (relative to the Midwest) was large, 917 pounds per cow, and statistically significant ($P < .01$). However, when the DHIA variable was added the effect dropped to 418 pounds and was not statistically significant ($P = .15$). When management was measured by the first two dimensions from correspondence analysis western region herds produced only 105 pounds less per cow than Midwestern herds and this difference was highly insignificant ($P = .73$). The addition of management also impacted the association of BST use and milk production per cow. In the base model each square root percentage increase in BST use was associated with a 363 pound increase in herd average milk production. When management was taken into consideration, the associated increase was not as great, 311 pounds for the model with DHIA variable and 242 pounds for the model with the first two correspondence analysis dimensions of the 24 management variables.

Using Dimensions of Management Practices to Classify Farms

Not only can correspondence analysis group various management practices by their dimension scores, it can group farms by dimension scores as well. Then the farm groupings can be evaluated for their association with disease and the economic impact associated with disease. An example of this can be shown from analysis of cow-calf data for the impact of Johne's disease on value of production (calf value at weaning + cull cow revenue – cow replacement cost).

In 1997 NAHMS surveyed 2,713 beef cow-calf producers across 23 states (USDA, 1997). A subset of these herds was tested for Johne's disease. One objective was to determine the economic impact of Johne's disease. An economic variable of value of production was calculated for each herd and expressed on a per cow basis. There were 363 herds that had 10 or more cows that were tested for Johne's disease. Twenty-eight of these herds tested positive with the disease. A problem with this study was that weaning weights were not actual scale weights but instead estimated weights by the producer. Over two-thirds of the estimated weaning weights were divisible by 50 and three-quarters were divisible by 25. Clearly, weaning weights were guesses and may have represented what the producer hoped his animals weighed and not an honest estimate of their actual weight. With weaned calf revenue being the largest component of value of production and it being a function of weaning weights, it should not be surprising that not many variables were associated with value of production. Only region (western vs. eastern), herd size (<50 cows vs. ≥50cows) and weight of the dam proved significant. Initial analysis of losses associated with Johne's disease found that Johne's test-positive herds had \$7 per cow less in value of production than Johne's test-negative herds and this difference was not statistically significant. (Because so few herds were in the analysis, the normal weighted analysis that

NAHMS does was not done and thus, SAS PROC REG (SAS Institute) was the regression algorithm used.)

Once again it was decided to add management ability via correspondence analysis. Twenty-nine management practices covering financial management (3 variables), feed management (7 variables), reproduction management (8 variables), vaccination use (6 variables) and use of expert information (5 variables) were incorporated into the analysis via correspondence analysis using SAS PROC CORRESP (SAS Institute). The first dimension captured 20 percent of the inertia of the 29 management variables, while the second dimension captured another 5 percent. Only the first dimension was statistically significant and its addition reduced the estimated loss associated with Johne's test-positive herds to \$4 per cow, which was also statistically insignificant.

An alternative approach was tried to investigate the economic impact of Johne's disease. The first dimension focused on the use of any of the management variables. The herds were divided into two categories based on whether their first dimension score was positive or negative. Of the 363 herds in this analysis, 202 had a positive first dimension score and the remaining 161 had a negative score. The herds with a positive first dimension used an average of 17.0 of the 29 management practices with a range of 11 to 25. The herds with a negative score used on average 8.6 management practices with a range of 2 to 15. Thus herds with a positive first dimension score were defined as being better managed herds while those with a negative score were defined as more poorly managed herds. Management practices that had the greatest influence in determining the herd's first dimension scores were balancing feed rations, using growth

promotants, various uses of veterinarians, computerized records, several reproductive management practices, and several vaccinations.

The first step was to test if there were any differences in the likelihood of being positive for Johne's disease. SAS PROC FREQ (SAS Institute) was used to compare percentages of herd infection. Within the better managed herds only 10 of the 202 (5%) herds tested positive for Johne's disease, but in the more poorly managed herds 18 of the 143 (11%) herds tested positive for Johne's disease; this difference was statistically significant ($P < .05$). Thus, more poorly managed herds were twice as likely to be test positive for Johne's disease.

The second step was to divide the herds into four groups based upon their Johne's disease status and management status. Then an ANOVA analysis was done using SAS PROC GLM (SAS Institute) to determine if there were any significant differences in value of production among the four groups. Herd size, location, and dam weight were included as covariates. Among better managed herds, the herds that tested positive for Johne's disease generated \$53 per cow less in value of production than Johne's test-negative herds and was statistically significant ($P < .05$) (table 3). On the other hand, for more poorly managed herds, which were more likely to be positive with Johne's disease, the disease did not negatively impact on value of production. In fact, value of production was greater on Johne's test-positive herds by \$25 per cow than on Johne's test-negative herds, but this amount was not statistically significant. Among the Johne's test-negative herds, better managed herds generated \$23 more per cow value of production than more poorly managed herds and this difference was statistically significant ($P < .05$). Thus, better managed herds generate more value of production than more poorly managed herds and

were less likely to be test-positive with Johne's disease. However, if they were test-positive with Johne's disease they take a financial hit, while the financial hit of Johne's disease among more poorly managed herds is nonexistent. A possible explanation of this result is that because of better management the better managed herds were more productive which is then reflected in their value of production. These better managers were pushing the productivity of their animals. Because they were pushing the productivity of their animals, when Johne's infection enters such a herd, productivity losses can be great. On the other hand, the more poorly managed herds which employed fewer of the management practices were not pushing the productivity limits of their herds. Also, they were not practicing preventive health measures as good as better managed herds and so they may have had more infection from other diseases. Their productivity was less and so when another disease entered their herds, it became one more disease and thus it didn't have much impact.

Summary

Quantifying management ability of farm producers has always been difficult. In the past researchers have used producer attributes, such as age and education, as proxy for management ability. Another way of measuring management ability is to assess the management practices employed by a producer. If several possible management practices are observed it would be helpful to have some method to combine them into a few "super" management variables. Such combination will eliminate potential multicollinearity problems that might exist if all the management practices were entered into an econometric model. Also, it is desirable to be able to classify producers objectively by the management practices they employed. A way to do all of

this is through correspondence analysis. Correspondence analysis is a statistical technique that can capture much of the information contained in several dichotomous variables into a couple of variables known as dimensions. The scores from the first couple of dimensions can be used in econometric models to represent several management practices. The more management can be accounted for, the more confidence we can have that the estimated association between variables of interest is indeed true and not because one or more of the variables serves as a proxy for management. This can be seen in the association between herd size and location with that of milk productivity from a dairy cow. When management was excluded from the analysis herd size and location were quite important in explaining the variance of milk production per cow. However, when management was incorporated into the analysis via correspondence analysis, these two variables become much less important in explaining milk productivity. In addition, management dimensions can be used to classify herds into better and poorer managed operations. This affords the analyst the ability to compare the association between management and herd health or profitability.

References

Dohoo, I.R., C. Ducrot, C. Fourichon, A. Donald, and D. Hurnik. "An Overview of Techniques for Dealing with Large Numbers of Independent Variables in Epidemiologic Studies." *Preventive Veterinary Medicine* 29(1996):221-239.

Greenacre, M.J. *Correspondence Analysis in Practice*. New York: Academic Press, 1993.

Greenacre, M.J. *Theory and Applications of Correspondence Analysis*. New York: Academic Press, 1984.

Hoffman, D.L. and G.R. Franke. "Correspondence analysis: Graphical representation of categorical data in marketing research." *J. Marketing Research* 23(1986):213-227.

Ott, S.L. and P.R. Novak. "Association of Herd Productivity and Bulk-tank Somatic Cell Counts in US Dairy Herds in 1996." *J. American Veterinary Medicine Association* 218(2001):1325-1330.

Ott, S.L., S.J. Wells, and B.A. Wagner. "Herd-level Economic Losses Associated with Johne's Disease on US Dairy Operations." *Preventive Veterinary Medicine* 40(1999):179-192.

SAS Institute. *SAS/SAT® User's Guide*, Version 8, Cary, NC: SAS Institute. 1999.

Shah, B.V, B.G. Barnwell, and G.S. Bieler. *SUDAAN User's Manual*, Release 7.5. Research Triangle Park, NC: Research

U.S. Department of Agriculture, Animal and Plant Health Inspection Service. *Part I: Reference of 1996 Dairy Management Practices*. USDA:APHIS:VS, Centers for Epidemiology and Animal Health, Fort Collins, CO May, 1996.

U.S. Department of Agriculture, Animal and Plant Health Inspection Service. *Part III: Reference of 1996 Dairy Health and Management Practices*. USDA:APHIS:VS, Centers for Epidemiology and Animal Health, Fort Collins, CO November, 1996.

U.S. Department of Agriculture, Animal and Plant Health Inspection Service. *Part I: Reference of 1997 Beef Cow-Calf Management Practices*. USDA:APHIS:VS, Centers for Epidemiology and Animal Health, Fort Collins, CO June, 1997.

Table 1. Comparison Among Models That Account for Management Practices in Explaining Variance of Herd-Level Milk Production Per Cow Among U.S. Dairies.

Explanatory Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Base Variables ^a	yes	yes	yes	yes	yes
Dairy Herd Improvement Assoc.**		yes			
Management index**			yes		
Dimension 1**				yes	yes
Dimension 2**					yes
Model R ²	0.409	0.473	0.469	0.457	0.499

a. Three levels of bulk tank somatic cell count, herd size, 4 levels of region, intensive pasture use, bovine somatotropin (BST) use, average days dry ≥ 70 days, and percent of herd being Holstein cattle. In Model 1 all variables were statistically significant at $P < .01$, except for days dry which was at $P < .05$.

** statistically significant at $P < .01$.

Table 2. Impact of Including Management as an Explanatory Variable on the Association of Base Model Explanatory Variables with Milk Production per Cow

Explanatory Variables	Base Model	Base Model with Dairy Herd Improvement Association Variable	Base Model with 1 st & 2 nd Dimensions of 24 Management Variables
Intercept	9827** ^a (819) ^b	9361** (739)	11823** (743)
Somatic cell count (200k – 399k/ml)	-993** (235)	-853** (214)	-782** (221)
Somatic cell count (400k + /ml)	-2577** (310)	-2194** (302)	-1977** (305)
Herd size (log of cow numbers)	1002** (168)	796** (154)	484** (161)
Western region (relative to Midwest region)	-917** (317)	-418 (296)	-105 (308)
Southeast region (relative to Midwest region)	-2959** (449)	-2475** (422)	-2492** (372)
Northeast region (relative to Midwest region)	-686** (256)	-513* (241)	-254 (245)
Intensive Pasture Use (pasture provides 90% of summer forage)	-1424** (299)	-1054** (279)	-1013** (286)
Bovine Somatotropin use (square root of % cows receiving bST)	363** (43)	311** (37)	242** (37)
Average days cows dry, $\geq 70d$	-672* (287)	-706** (273)	-688** (257)
Percent of herd Holstein cattle	51** (5)	52** (4)	49** (4)
DHIA use		1915** (201)	
Dimension 1 ^c			-2078** (241)
Dimension 2 ^d			2738** (338)
Model R ²	0.409	0.473	0.499

a. Pounds of milk per cow.

b. Standard error.

c. The more management practices used, the less the dimension score.

d. The more vaccinations used, the greater the dimension score.

**Statistically significance, $P < .01$.

* Statistically significance, $P < .05$.

Table 3. Value of Production¹ by Management and Johne's Disease Status in Beef Cow Herds (\$ per Cow)²

	Johne's Test-Positive Herds	Johne's Test-Negative Herds
Better Managed Herds ³	\$209 ^{4,a,c} n = 10 row percentage = 5.0% ^e	\$262 ^{a,d} n = 192 row percentage = 95.0% ^f
Poorer Managed Herds	\$264 ^{b,c} n = 18 row percentage = 11.2% ^e	\$239 ^{b,d} n = 143 row percentage = 88.8% ^f

1. Value of production = calf value at weaning + cull cow revenue – cow replacement cost.
2. Compared using ANOVA analysis with herd size, location, and weight of dam as covariates.
3. Better and poorer managed herds determined by first dimension score from correspondence analysis. The first dimension focused on overall use of various management practices. Herds with a positive score were defined as better managed herds and employed more management practices than herds with a negative score or the poorer managed herds.
4. Values with the same superscript were compared statistically and their level of significance are listed below
 - a. P = .022
 - b. P = .155
 - c. P = .050
 - d. P = .010
 - e. P < .05
 - f. P < .05

Appendix 1. An Example Correspondence Analysis Code in SAS.

```
/* read in data file */
data temp1; set sasdata.datafile;

/* correspondence analysis with 11 management variables */
/* output the 10 dimensions to a new data set */
/* suppress printing */
proc corresp outc=newdata noprint;
  var m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11;

/* read in calculated dimension values */
data temp2; set newdata;

/* for variable reduction to use in regression analysis delete
   _TYPE_ = INERTIA and _TYPE_ = VAR */
if _TYPE_ = 'INERTIA' then delete;
if _TYPE_ = 'VAR' then delete;

keep dim1 dim2 dim3 dim4 dim5 dim6 dim7 dim8 dim9 dim10 _NAME_;

/* merge dimension data set with original data set */
data temp3; merge temp1 temp2;

/* regression analysis */
proc reg;
  model depvar = var1 var2 dim1 dim2;
run;

/* NOTE: If one desires to group the management variables by dimension scores then delete
   _TYPE_ = 'INERTIA' and _TYPE_ = 'OBS'. The management variables will be listed under
   _NAME_ */
/* example of grouping management variables into 4 quadrants formed by the zero line of the
   first two dimensions */
data temp4; set newdata;
if _TYPE_ = 'INERTIA' then delete;
if _TYPE_ = 'OBS' then delete;

/* form the four quadrants */
if dim1 >= 0 and dim2 >= 0 then quad = '++';
if dim1 < 0 and dim2 >= 0 then quad = '-+';
if dim1 >= 0 and dim2 < 0 then quad = '+-';
if dim1 < 0 and dim2 < 0 then quad = '--';

proc sort; by quad;
proc print;
```

```
var _NAME;  
by quad;  
run;
```