Journal of Modern Applied Statistical Methods

Volume 4 | Issue 1 Article 7

5-1-2005

Regression By Data Segments Via Discriminant Analysis

Stan Lipovetsky
GfK Custom Research North America, Minneapolis, MN, stan.lipovetsky@gfk.com

Michael Conklin
GfK Custom Research Inc

Journal of Modern Applied Statistical Methods
May, 2005, Vol. 4, No.1, 63-74

1538 – 9472/05/$95.00

Regression By Data Segments Via Discriminant Analysis

Stan Lipovetsky Michael Conklin
GfK Custom Research Inc
Minneapolis, Minnesota

It is known that two-group linear discriminant function can be constructed via binary regression. In this article, it is shown that the opposite relation is also relevant – it is possible to present multiple regression as a linear combination of a main part, based on the pooled variance, and Fisher discriminators by data segments. Presenting regression as an aggregate of the discriminators allows one to decompose coefficients of the model into sum of several vectors related to segments. Using this technique provides an understanding of how the total regression model is composed of the regressions by the segments with possible opposite directions of the dependency on the predictors.

Key words: Regression, discriminant analysis, data segments

Introduction

Linear Discriminant Analysis (LDA) was introduced by Fisher (1936) for classification of observations into two groups by maximizing the ratio of between-group variance to within-group variance (Rao, 1973; Lachenbruch, 1979; Hand, 1982; Dillon & Goldstein, 1984; McLachlan, 1992; Huberty, 1994). For two-group LDA, the Fisher linear discriminant function can be represented as a linear regression of a binary variable (groups indicator) by the predictors (Fisher, 1936; Anderson, 1958; Ladd, 1966; Hastie, Tibshirani & Buja, 1994; Ripley, 1996). Many-group LDA can be described in terms of the Canonical Correlations Analysis (Bartlett, 1938; Kendall & Stuart, 1966; Dillon & Goldstein, 1984; Lipovetsky, Tishler, & Conklin, 2002). LDA is used in various applications, for example, in marketing research (Morrison, 1974; Hora & Wilcox, 1982; Lipovetsky & Conklin, 2004).

Considered in this article is the possibility of presenting a multiple regression by segmented data as a linear combination of the Fisher discriminant functions. This technique is based on the relationship between total and pooled variances. Using this approach, we can interpret regression as an aggregate of discriminators, that allows us to decompose the coefficients of regression into a sum of vectors related to the data segments. Such a decomposition helps explain how a regression by total data could have the opposite direction of the dependency on the predictors, in comparison with the coefficients related to each segment.

These effects correspond to well-known Simpson's and Lord's paradoxes (Blyth, 1972; Holland & Rubin, 1983; Good & Mittal, 1987; Pearl, 2000; Rinott & Tam, 2003; Skrondal & Rabe-Hesketh, 2004; Wainer & Brown, 2004), and to treatment and causal effects in the models (Arminger, Clogg & Sobel, 1995; Rosenbaum, 1995; Winship & Morgan, 1999).

The article is organized as follows. Linear discriminant analysis and its relation to binary regression are first described. The next section considers regression by segmented data and its decomposition by Fisher discriminators, followed by a numerical example and a summary.

Stan Lipovetsky joined GfK Custom Research as a Research Manager in 1998. His primary areas of research are multivariate statistics, multiple criteria decision making, econometrics, and marketing research. Email him at: slipovetsky@gfkcustomresearch.com. Michael Conklin is Senior Vice-President, Analytic Services for GfK Custom Research. His research interests include Bayesian methods and the analysis of categorical and ordinal data.

Methodology

Consider the main features of LDA. Denote X a data matrix of n by p order consisting of n rows of observations by p variables $x_1, x_2, \ldots, x_p$. Also denote y a vector of size n consisting of binary values 1 or 0 that indicate belonging of each observations to one or another class. Suppose there are n_1 observations in the first class (y =1), n_2 observations in the second class (y =0), and total number of observations $n=n_1+n_2$. Construct a linear aggregate of x-variables:

$$z = X\,a\,, \tag{1}$$

where a is a vector of p-th order of unknown parameters, and z is an n-th order vector of the aggregate scores. Averaging scores z (1) within each group yields two aggregates:

$$z^{(1)} = m^{(1)}a\,, \quad z^{(2)} = m^{(2)}a\,, \tag{2}$$

where $m^{(1)}$ and $m^{(2)}$ are vectors of p-th order of mean values $m_j^{(1)}$ and $m_j^{(2)}$ of each j-th variable x_j within the first and second group of observations, respectively. The maximum squared distance between two groups $\|z^{(1)}-z^{(2)}\|^2 = \|(m^{(1)}-m^{(2)})a\|^2$ versus the pooled variance of scores $a'S_{pool}\ a$ defines the objective for linear discriminator:

$$F = \frac{a'(m^{(1)} - m^{(2)})(m^{(1)} - m^{(2)})'a}{a'S_{pool}\ a}\,, \tag{3}$$

with elements of the pooled matrix defined by combined cross-products of both groups:

$$(S_{pool})_{jk} = \sum_{i=1}^{n_1}(x_{ji} - m_j^{(1)})(x_{ki} - m_k^{(1)}) + \sum_{i=1}^{n_2}(x_{ji} - m_j^{(2)})(x_{ki} - m_k^{(2)})\,. \tag{4}$$

Equation (3) can represent as a conditional objective:

$$F = a'(m^{(1)} - m^{(2)})(m^{(1)} - m^{(2)})'a - \lambda(a'S_{pool}\ a - 1)\,, \tag{5}$$

where λ is Lagrange multiplier. The first-order condition $\partial F/\partial a = 0$ yields:

$$(m^{(1)} - m^{(2)})(m^{(1)} - m^{(2)})'a = \lambda\, S_{pool}\ a\,, \tag{6}$$

that is a generalized eigenproblem. The matrix at the left-hand side (6) is of the rank one because it equals the outer product of a vector of the group means' differences. So the problem (6) has just one eigenvalue different from zero and can be simplified. Using a constant of the scalar product $c = (m^{(1)} - m^{(2)})'a$, reduces (6) to the linear system:

$$S_{pool}\ a = q\,(m^{(1)} - m^{(2)})\,, \tag{7}$$

where $q=c/\lambda$ is another constant. The solution of this system is:

$$a = S_{pool}^{-1}\,(m^{(1)} - m^{(2)})\,, \tag{8}$$

that defines Fisher famous two-group linear discriminator (up to an arbitrary constant).

The same Fisher discriminator (8) can be obtained if instead of the pooled matrix (4) the total matrix of second-moments defined as a cross-product $X'X$ of the centered data is used, so the elements of this matrix are:

$$(S_{tot})_{jk} = \sum_{i=1}^{n}(x_{ji} - m_j)(x_{ki} - m_k)\,, \tag{9}$$

where m_j corresponds to mean value of each x_j by total sample of size n. Similarly to transformation known in the analysis of variance, consider decomposition of the cross-product (9) into several items when the total set of n observations is divided into subsets with sizes n_t with t = 1, 2, …, T :

$$\begin{aligned}(S_{tot})_{jk} &= \sum_{i=1}^{n}(x_{ji} - m_j)(x_{ki} - m_k) = \sum_{t=1}^{T}\sum_{i=1}^{n_t}[(x_{ji}^{(t)} - m_j^{(t)}) + (m_j^{(t)} - m_j)]x_{ki}^{(t)} \\ &= \sum_{t=1}^{T}\sum_{i=1}^{n_t}(x_{ji}^{(t)} - m_j^{(t)})x_{ki}^{(t)} + \sum_{t=1}^{T}\sum_{i=1}^{n_t}(m_j^{(t)} - m_j)x_{ki}^{(t)} \\ &= \sum_{t=1}^{T}\sum_{i=1}^{n_t}(x_{ji}^{(t)} - m_j^{(t)})(x_{ki}^{(t)} - m_k^{(t)}) + \sum_{t=1}^{T} n_t\,(m_j^{(t)} - m_j)(m_k^{(t)} - m_k)\,.\end{aligned} \tag{10}$$

The obtained double sum equals the pooled second moment (4) for T groups, and the last sum corresponds to a total (weighted by sub-sample sizes) of the second moment of group means centered by the total means of the variables. So (10) can be rewrote in a matrix form as:

$$S_{tot} = S_{pool} + \sum_{t=1}^{T} n_t (m^{(t)} - m)(m^{(t)} - m)', \quad (11)$$

where $m^{(t)}$ is a vector of mean values $m_j^{(t)}$ of each j-th variable within t-th group, and m is a vector of means for all variables by the total sample.

Consider the case of two groups, T=2. Then (11) can be reduced to

$$\begin{aligned} S_{tot} &= S_{pool} + n_1 \left(m^{(1)} - \frac{n_1 m^{(1)} + n_2 m^{(2)}}{n_1 + n_2} \right) \\ &\quad \cdot \left(m^{(1)} - \frac{n_1 m^{(1)} + n_2 m^{(2)}}{n_1 + n_2} \right)' \\ &\quad + n_2 \left(m^{(2)} - \frac{n_1 m^{(1)} + n_2 m^{(2)}}{n_1 + n_2} \right) \\ &\quad \cdot \left(m^{(2)} - \frac{n_1 m^{(1)} + n_2 m^{(2)}}{n_1 + n_2} \right)' \\ &= S_{pool} + h (m^{(1)} - m^{(2)})(m^{(1)} - m^{(2)})', \end{aligned} \quad (12)$$

where $h = n_1 n_2 / (n_1 + n_2)$ is a constant of the harmonic sum of sub-sample sizes. In place of the pooled matrix S_{pool} let us use the total matrix S_{tot} (12) in the LDA problem (7):

$$\begin{aligned} &\left(S_{pool} + h (m^{(1)} - m^{(2)})(m^{(1)} - m^{(2)})' \right) a \\ &= q (m^{(1)} - m^{(2)}) \end{aligned} . \quad (13)$$

Applying a known Sherman-Morrison formula (Rao, 1973; Harville, 1997)

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1} u v' A^{-1}}{1 + u' A^{-1} v}, \quad (14)$$

where A is a non-singular square n-th order matrix, u and v are vectors of n-th order, the matrix in the left-hand side (13) is inverted and solution obtained:

$$\begin{aligned} a &= S_{tot}^{-1} (m^{(1)} - m^{(2)}) q \\ &= \frac{q}{1 + h (m^{(1)} - m^{(2)})' S_{pool}^{-1} (m^{(1)} - m^{(2)})} \\ &\quad \cdot S_{pool}^{-1} (m^{(1)} - m^{(2)}). \end{aligned} \quad (15)$$

Comparison of (8) and (15) shows that both discriminant functions coincide (up to unimportant in LDA constant in the denominator (15)), so we can use S_{tot} instead of S_{pool} .

This feature of proportional solutions for the pooled or total matrices holds for more than two classification groups as well. Consider a criterion of maximizing ratio (3) of between-group to the within-group variances for many groups. Using the relation (11) yields:

$$\begin{aligned} F &= \frac{a' (S_{tot} - S_{pool}) a}{a' S_{pool} a} \\ &= \frac{a' \left(\sum_{t=1}^{T} n_t (m^{(t)} - m)(m^{(t)} - m)' \right) a}{a' S_{pool} a} \end{aligned} . \quad (16)$$

Similarly to derivation (5)-(6), (16) is reduced to an eigenproblem:

$$\left(\sum_{t=1}^{T} n_t (m^{(t)} - m)(m^{(t)} - m)' \right) a = \lambda S_{pool} a, \quad (17)$$

that is a generalized eigenproblem for the many groups. Denoting the scalar products at the left-hand side (17) as some constants $c_t = (m^{(t)} - m)' a$, the solution of (22) via a linear combination of Fisher discriminators is presented:

$$a = \sum_{t=1}^{T} c_t n_t S_{pool}^{-1} (m^{(t)} - m). \quad (18)$$

In the case of two groups we have simplification (12) that reduces the eigenproblem (17) to the solution (8). But the discriminant functions in multi-group LDA with the pooled matrix or the total matrix in (17) are the same (up to a normalization) – a feature similar to two group LDA (15). To show this, rewrite (17) using (16) in terms of these two matrices as a generalized eigenproblem:

$$(S_{tot} - S_{pool})a = \lambda S_{pool}\, a\,. \qquad (19)$$

Multiplying S_{pool}^{-1} by the relation (19) reduces it to a regular eigenproblem $(S_{pool}^{-1} S_{tot})\, a = (\lambda + 1)\, a$. Taking the objective (16) with the total matrix in denominator, another generalized eigenproblem is obtained:

$$(S_{tot} - S_{pool})b = \mu S_{tot}\, b\,, \qquad (20)$$

with eigenvalues μ and eigenvectors b in this case. Multiplying S_{pool}^{-1} by the relation (20), it is represented as $(S_{pool}^{-1} S_{tot})\, b = (1/(1-\mu))\, b$. Both problems (19) and (20) are reduced to the eigenproblem for the same matrix $S_{pool}^{-1} S_{tot}$ with the eigenvalues connected as $(1+\lambda)(1-\mu)=1$ and with the coinciding eigenvectors a and b.

Now, consider some properties of linear regression related to discriminant analysis. Multiple regression can be presented in a matrix form as a model:

$$y = X\, a + \varepsilon\,, \qquad (21)$$

where Xa is a vector of theoretical values of the dependent variable y (corresponding to the linear aggregate z (1)), and ε denotes a vector of errors. The Least Squares objective for minimizing is:

$$\begin{aligned} LS &= \|\varepsilon\|^2 = (y - Xa)'(y - Xa) \\ &= y'y - 2a'X'\,y + a'X'Xa \end{aligned}\,. \qquad (22)$$

The condition for minimization $\partial LS / \partial a = 0$ yields a normal system of equations:

$$(X'X)a = X'y\,, \qquad (23)$$

with the solution for the coefficients of the regression model:

$$a = (X'X)^{-1} X'y\,. \qquad (24)$$

Matrix of the second moments $X'X$ in (23) for the centered data is the same matrix S_{tot} (9). If the dependent variable y is binary, then the vector $X'y$ is proportional to the vector of differences between mean values by two groups $m^{(1)} - m^{(2)}$, and solution (24) is proportional to the solution (15) for the discriminant function defined via S_{tot}. As it was shown in (15), the results of LDA are essentially the same with both S_{tot} or S_{pool} matrices. Although the Fisher discriminator can be obtained in regular linear regression of the binary group indicator variable by the predictors, a linear regression with binary output can also be interpreted as a Fisher discriminator. Predictions $z=Xa$ (21) by the regression model are proportional to the classification (1) by the discriminator (15).

Regression as an Aggregate of Discriminators

Now, the regression is described by data segments presented via an aggregate of discriminators. Suppose the data are segmented; for instance, the segments are defined by clustering the independent variables, or by several intervals within a span of the dependent variable variation. Identify the segments by index t =1,…,T to present the total second-moment matrix $S_{tot} = X'X$ as the sum (11) of the pooled second-moment matrix S_{pool} and the total of outer products for the vectors of deviations of each segment's means from the total means. Using the relation (11), the normal system of equations (23) for linear regression is represented as follows:

$$\left(S_{pool} + \sum_{t=1}^{T} n_t (m^{(t)} - m)(m^{(t)} - m)' \right) a = X'y\,. \qquad (25)$$

where the pooled cross-product is defined due to (10)-(11) as:

$$S_{pool} = \sum_{t=1}^{T}\sum_{i=1}^{n_t}(x_{ji}^{(t)} - m_j^{(t)})(x_{ki}^{(t)} - m_k^{(t)}) \equiv \sum_{t=1}^{T} S_t \,, \tag{26}$$

where S_t are the matrices of second moments within each t-th segment. Introducing the constants

$$c_t = (m^{(t)} - m)'a \,, \tag{27}$$

defined similarly to those in derivation (17)-(18), reducing the system (25) to:

$$S_{pool}\, a = X'y - \sum_{t=1}^{T} n_t c_t \,(m^{(t)} - m) \,. \tag{28}$$

Then solution of (28) is:

$$\begin{aligned} a &= S_{pool}^{-1}\, X'y - \sum_{t=1}^{T} n_t c_t \, S_{pool}^{-1}(m^{(t)} - m) \\ &\equiv a_{pool} - \sum_{t=1}^{T} n_t c_t \, a_t \end{aligned} \,. \tag{29}$$

In (29) the notations used are:

$$a_{pool} = S_{pool}^{-1}\, X'y \,, \quad a_t = S_{pool}^{-1}(m^{(t)} - m) \,, \tag{30}$$

so the vector a_{pool} corresponds to the main part of the total vector in (29) of the regression coefficients defined via the pooled matrix (26), and additional vectors a_t correspond to Fisher discriminators (8) between each t-th particular segment and total data set. Decomposition (29) shows that regression coefficients a consist of the part a_{pool} and a linear aggregate (with weights $n_t c_t$) of Fisher discriminators a_t of the segments versus total data. It is interesting to note that if to increase number of segments up to the number of observations ($T=n$, with only one observation in each segment) then each variable's mean in any segment coincides with the original observation itself, $m_k^{(t)} = x_{ki}^{(t)}$, so $S_{pool} = 0$ in (26). In this case the sum in (25) coincides with the total second-moment matrix, so the regular regression solution can be seen as an aggregate of the discriminators by each observation versus total vector of means.

The obtained decomposition (29) is useful for interpretation, but it still contains the unknown parameters c_t (27) that need to be estimated. First, notice that the Fisher discriminators a_t (30) of each segment versus entire data, are restricted by the relation:

$$\begin{aligned} \sum_{t=1}^{T} n_t\, a_t &= \sum_{t=1}^{T} n_t S_{pool}^{-1}(m^{(t)} - m) \\ &= S_{pool}^{-1}\sum_{t=1}^{T} n_t (m^{(t)} - m) \\ &= S_{pool}^{-1}\left(\sum_{t=1}^{T} n_t m^{(t)} - m\sum_{t=1}^{T} n_t\right) = 0 \end{aligned} \,. \tag{31}$$

Thus, for T segments there are only T-1 independent discriminators.

Consider a simple case of two segments in data. In difference to the described two-group LDA problem (12)-(15) and its relation to the binary linear regression (24), we can have a non-binary output, for instance, a continuous dependent variable. Using the derivation (12)-(15) for the inversion of the matrix of the normal system of equations (25), the solution (29) is obtained for two-segment linear regression in explicit form:

$$\begin{aligned} a &= S_{tot}^{-1}X'y \\ &= \left(S_{pool}^{-1} - \frac{hS_{pool}^{-1}(m^{(1)} - m^{(2)})(m^{(1)} - m^{(2)})'S_{pool}^{-1}}{1 + h(m^{(1)} - m^{(2)})'S_{pool}^{-1}(m^{(1)} - m^{(2)})} \right) X'y \\ &= S_{pool}^{-1}X'y - \left(\frac{h(m^{(1)} - m^{(2)})'S_{pool}^{-1}X'y}{1 + h(m^{(1)} - m^{(2)})'S_{pool}^{-1}(m^{(1)} - m^{(2)})} \right) \\ &\quad \cdot S_{pool}^{-1}(m^{(1)} - m^{(2)}) \,, \end{aligned} \tag{32}$$

where h is the same constant as in (12). It can be seen that the vector of coefficients for two-segment regression, similarly to the general solution (29), equals the main part a_{pool} (30) minus a constant (in the parentheses at the right-hand side (32) multiplied by the discriminator (8).

Another analytical result can be obtained for three segments in data, when a general solution (29) contains two discriminators. For this case we extended the Sherman-Morrison formula (14) to the inversion of a matrix $A+u_1v_1'+u_2v_2'$, where A is a non-singular matrix and $u_1v_1'+u_2v_2'$ are two outer products of vectors. The derivation for the inverted matrix of such a structure is given in the Appendix. In this case, the system (25) can be presented in the notations:

$$A=S_{pool}\,,\quad u_1=v_1=\sqrt{n_1}(m^{(1)}-m)\,,\quad u_2=v_2=\sqrt{n_2}(m^{(2)}-m) \qquad (33)$$

Applying the formula (A16) with definitions (33), we obtain solution of the system (25) for three segments. In accordance with the relations (29)-(31), this solution is expressed via the vector a_{pool} and two Fisher discriminators.

In a general case of any number T of segments, the parameters c_t in the decomposition (29) can be obtained in the following procedure. Theoretical values of the dependent variable are predicted by the regression model (28) as follows:

$$\tilde{y}=X\,a=XS_{pool}^{-1}\,X'y+\sum_{t=1}^{T-1}c_t[n_tX\,S_{pool}^{-1}(m-m^{(t)})]\equiv\tilde{y}_{pool}+\sum_{t=1}^{T-1}c_t\tilde{y}_t\,, \qquad (34)$$

where a predicted vector $\tilde{y}$ is decomposed to the vector $\tilde{y}_{pool}$ defined via the pooled variance and the items $\tilde{y}_t$ related to the Fisher discriminator functions in the prediction:

$$\tilde{y}_{pool}=XS_{pool}^{-1}\,X'y\,,\quad \tilde{y}_t=n_tXS_{pool}^{-1}(m-m^{(t)})\,. \qquad (35)$$

All the vectors in (35) can be found from the data, so using $\tilde{y}$ (34) in the regression (21), the model is reduced to:

$$\Delta y=\sum_{t=1}^{T-1}c_t\tilde{y}_t+\varepsilon\,, \qquad (36)$$

where $\Delta y=y-\tilde{y}_{pool}$ is a vector of difference between empirical and predicted by pooled variance theoretical values of the dependent variable. The relation (36) is also a model of regression of the dependent variable Δy by the new predictors - the Fisher classifications $\tilde{y}_t$ (35). This regression can be constructed in the Least Squares approach (22)-(24). In difference to the regression (21) by possibly many independent x variables, the model (36) contains just a few regressors $\tilde{y}_t$, because a number of segments is usually small.

Regression decomposition (25)-(35) uses the segments within the independent variables, that is expressed in presentation of the total second-moment matrix of x-s at the left-hand side (25) via the pooled matrix of x-s (26). However, there is also a vector $X'y$ of the x-s cross-products with the dependent variable y at the right-hand side of normal system of equations (25). The decomposition of this vector can also be performed by the relations (10)-(11). Suppose, we use the same segments for all x-s and y variables, then:

$$X'y\equiv(X'y)_{tot}=(X'y)_{pool}+\sum_{t=1}^{T}n_t(m^{(t)}-m)(\bar{y}^{(t)}-\bar{y})\,, \qquad (37)$$

where $\bar{y}^{(t)}$ and $\bar{y}$ are the mean values of the dependent variable in each t-th segment and the total mean. The elements of the vector $(X'y)_{pool}$ in (37) are defined due to (10)-(11) as:

$$(x_j'y)_{pool}=\sum_{t=1}^{T}\sum_{i=1}^{n_t}(x_{ji}^{(t)}-m_j^{(t)})(y_i^{(t)}-\bar{y}^{(t)})\,, \qquad (38)$$

where x_j is a column of observations for the j-th variable in the X matrix. Using the presentation (37)-(38) in place of the vector $X'y$ in (29)-(30) yields a more detailed decomposition of the vector a_{pool} by the segments within the dependent variable data. In the other relations (32), or (34)-(35), this further decomposition can be used as well. In a more general case we can consider different segments for the independent variables and for the dependent variable y.

If y is an ordinal variable, and the segments are chosen by its levels, then within each segment there are zero equaled deviations $y_i^{(t)} - \bar{y}^{(t)} = 0$. Thus, in (38) the values $(x_j' y)_{pool} = 0$, and the decomposition (37) does not contain the pooled vector $(X'y)_{pool}$. Solution (29) can then be given as:

$$a = S_{pool}^{-1} \begin{pmatrix} \sum_{t=1}^{T} n_t (m^{(t)} - m)(\bar{y}^{(t)} - \bar{y}) \\ -\sum_{t=1}^{T} n_t c_t (m^{(t)} - m) \end{pmatrix} = \sum_{t=1}^{T} \gamma_t a_t \, , \tag{39}$$

where the vectors by segments and the constants are defined as:

$$a_t = S_{pool}^{-1} (m^{(t)} - m) \, , \quad \gamma_t = n_t (\bar{y}^{(t)} - \bar{y} - c_t) \, . \tag{40}$$

Thus, the solution (29)-(30) is in this case reduced to the linear combination of discriminant functions a_t with the weights γ_t, without the a_{pool} input. This solution corresponds to the classification (18) by several groups in discriminant analysis. The parameters γ_t can be estimated as it is described in the procedure (32)-(36). If we work with a centered data, a vector of total means by x-variables $m = 0$ and the mean value $\bar{y} = 0$, so these items can be omitted in all the formulae.

A useful property of the solution (30) consists in the inversion of the pooled matrix S_{pool} instead of inversion of the total matrix $S_{tot} = X'X$ as in (24). If the independent variables are multicollinear, their covariance or correlation matrix is ill-conditioned or close to a singular matrix. The condition number, defined as ratio between the biggest and the smallest eigenvalues, is large for the ill-conditioned matrices and even infinite for a singular matrix. For such a total matrix $X'X$ there could be a problem with its inversion. At the same time the pooled matrix obtained as a sum of segmented matrices (26), is usually less ill-conditioned. The numerical simulations showed that the condition numbers of the pooled matrices are regularly many times less than these values of the related total second-moment matrices. It means that working with a pooled matrix in (30) yields more robust results, not as prone to multicollinearity effects as in a regular regression approach.

Numerical example

Consider an example from a real research project with 550 observations, where the dependent variable is customer overall satisfaction with a bank merchant's services, and the independent variables are: x_1 – satisfaction with the account set up; x_2 – satisfaction with communication; x_3 – satisfaction with how sales representatives answer questions; x_4 – satisfaction with information needed for account application; x_5 – satisfaction with the account features; x_6 – satisfaction with rates and fees; x_7 – satisfaction with time to deposit into account. All variables are measured with a ten-point scale from absolutely non-satisfied to absolutely satisfied (1 to 10 values). The pair correlations of all variables are positive. The data is considered in three segments of non-satisfied, neutral, and definitely satisfied customers, where the segments correspond to the values of the dependent variable from 1 to 5, from 6 to 9, and 10, respectively.

Consider the segments' contribution into the regression coefficients and into the total model quality. The coefficients of regression for the standardized variables are presented in the last column of Table 1.

The coefficient of multiple determination for this model is R^2=0.485, and F-statistics equals 73.3, so the quality of the regression is good. The first four columns in Table 1 present inputs to the coefficients of regression from the pooled variance of the independent variables combined with the pooled variance of the dependent variable and three segments (37)-(38). The sum of these items in the next column comprises the pooled subtotal a_{pool} (30).

Table 1. Regression Decomposition by the Items of Pooled Variance and Discriminators.

Variable	Pooled Dependent	Pooled Variance of Predictors: Segment 1	Segment 2	Segment 3	Pooled Subtotal	Fisher Discriminators: Segment 1	Segment 3	Regression Total
x_1	.116	.026	.015	.064	.222	-.011	-.044	.166
x_2	.007	.149	.001	.049	.206	-.064	-.034	.108
x_3	.008	.232	-.006	.048	.282	-.100	-.033	.149
x_4	-.035	.005	.021	.077	.068	-.002	-.053	.013
x_5	.039	.101	-.016	-.028	.096	-.044	.019	.072
x_6	.054	.325	.012	.142	.533	-.141	-.098	.294
x_7	.048	.102	.018	.095	.262	-.044	-.065	.153

Table 2. Regression Decomposition by Segments.

Variable	Core Input: Coefficient	Net Effect	Segment 1: Coefficient	Net Effect	Segment3: Coefficient	Net Effect	Regression Total: Coefficient	Net Effect
x_1	.131	.072	.015	.008	.020	.011	.166	.091
x_2	.008	.005	.084	.046	.015	.008	.108	.059
x_3	.003	.001	.131	.069	.015	.008	.149	.078
x_4	-.014	-.006	.003	.001	.024	.011	.013	.006
x_5	.023	.008	.057	.020	-.009	-.003	.072	.025
x_6	.066	.037	.184	.103	.044	.025	.294	.165
x_7	.065	.026	.058	.023	.030	.012	.153	.061
R^2	.143		.271		.071		.485	
R^2 share	29%		56%		15%		100%	

The next two columns present the Fisher discriminators (30) for the first and the third segments. It is interesting to note that the condition numbers of the predictors total and pooled matrices of second moments equal 19.7 and 11.9, so the latter one is much less ill-conditioned. Adding the pooled subtotal a_{pool} and Fisher discriminators yields the total coefficients of regression in the last column of Table 1.

Combining some columns of the first table, Table 2 of the main contributions to the coefficients of regression is obtained. Table 2 consists of doubled columns containing coefficients of regression and the corresponded net effects. In Table 2, the core input coefficients equal the sum of pooled dependent and the segment-2 columns from Table 1. Segment-1 coefficients in Table 2 equal the sum of two columns related to Segment-1 from Table 1, and similarly for the Segment-3 coefficients.

Summing all three of these columns of coefficients in Table 2 yields the total coefficients of regression. Considering coefficients in the columns of Table 2 in a way similar to factor loadings in factor analysis, we can identify which variables are more important in each segment of the total coefficients of regression. For instance, comparing coefficients in each row across three first columns in Table 2, we see that the variables x_1 and x_7 have the bigger values in the core input than in segments, satisfaction with account set up and with time to deposit into account play a basic role in the customer overall satisfaction.

Segment-1 has bigger coefficients by the variables x_2, x_3, x_5, and x_6, and the Segment-3 has a bigger coefficient by the variable x_4, so the corresponded attributes play the major roles in creating customers dissatisfaction or delight, respectively. It is interesting to note that this approach produces similar results to another technique developed specifically for the customer satisfaction studies (Conklin, Powaga & Lipovetsky, 2004).

Besides the coefficients of regression, Table 2 presents the net effects, or the characteristics of comparative influence of the regressors in the model (for more on this topic, see Lipovetsky &Conklin, 2001). Quality of regression can be estimated by the coefficient of multiple determination defined by the scalar product of the standardized coefficients of regression a_j and the vector of pair correlations r_{yj} of the dependent variable and each j-th independent variables, so $r_{yj}=(X'y)_j$. Items $r_{yj}a_j$ in total R^2 are called the net effects of each predictor: $R^2 = r_{y1}a_1 + r_{y2}a_2 + ... r_{yn}a_n$. The net effects for core, two segment items, and their total (that is equal to the net effects obtained by the total coefficients of regression) are shown in Table 2.

The net effects can be also used for finding the important predictors in each component of total regression. Summing net effects within their columns in Table 2 yields a splitting of total R^2 =.485 into its core (R^2 =.143), segment-1 (R^2 =.271), and segment-3 (R^2 =.071) components. In the last row of Table 2 we see that the core and two segments contribute to total coefficient of multiple determination by 29%, 56%, and 15%, respectively. Thus, the main share in the regression is produced by segment-1 of the dissatisfaction influence.

Conclusion

Relations between linear discriminant analysis and multiple regression modeling were considered using decomposition of total matrix of second moments of predictors into pooled matrix and outer products of the vectors of segment means. It was demonstrated that regression coefficients can be presented as an aggregate of several items related to the pooled segments and Fisher discriminators. The relations between regression and discriminant analyses demonstrate how a total regression model is composed of the regressions by the segments with possible opposite directions of the dependency on the predictors. Using the suggested approach can provide a better understanding of regression properties and help to find an adequate interpretation of regression results.

References

Anderson T. W. (1958) *An introduction to multivariate statistical analysis*. New York: Wiley and Sons.

Arminger, G., Clogg C. C., & Sobel M. E., (Eds.) (1995). *Handbook of statistical modeling for the social and behavioral sciences.* London: Plenum Press.

Bartlett M. S. (1938). Farther aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, 34, 33-40.

Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67, 364-366.

Conklin, M., Powaga, K., & Lipovetsky, S. (2004). Customer satisfaction analysis: identification of key drivers. *European Journal of Operational Research, 154/3,* 819-827.

Dillon, W. R., & Goldstein, M. (1984) . *Multivariate analysis: methods and applications*. New York: Wiley and Sons.

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7,* 179-188.

Good, I. J., & Mittal, Y. (1987). The amalgamation and geometry of two-by-two contingency tables. *The Annals of Statistics, 15,* 694-711.

Hand, D. J. (1982). *Kernel discriminant analysis*. New York: Research Studies Press.

Hastie, T., Tibshirani, R., & Buja, A. (1994). Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association, 89*, 1255-1270.

Harville, D. A. (1997) *Matrix algebra from a statistician' perspective*. New York: Springer.

Holland, P. W., & Rubin D. B. (1983). On Lord's paradox. In: H. Wainer & S. Messick (Eds.) *Principals of Modern Psychological Measurement*, 3-25. Hillsdale, NJ: Lawrence Earlbaum.

Hora, S. C., & Wilcox, J. B. (1982). Estimation of error rates in several-population discriminant analysis. *Journal of Marketing Research*, *19*, 57-61.

Huberty, C. H. (1994). *Applied discriminant analysis*. New York: Wiley and Sons.

Kendall, M. G. & Stuart, A. (1966). *The advanced theory of statistics*, vol. III. London: Griffin.

Lachenbruch, P. A. (1979). Discriminant analysis. *Biometrics, 35,* 69-85.

Ladd, J. W. (1966) Linear probability functions and discriminant functions. *Econometrica*, 34, 873-885.

Lipovetsky, S., & Conklin M. (2001). Analysis of regression in game theory approach,. *Applied Stochastic Models in Business and Industry, 17,* 319-330.

Lipovetsky, S., & Conklin, M. (2004). Decision making by variable contribution in discriminant, logit, and regression analyses. *Information Technology and Decision Making, 3*, 265-279.

Lipovetsky, S., Tishler A., & Conklin. M. (2002). Multivariate least squares and its relation to other multivariate techniques. *Applied Stochastic Models in Business and Industry, 18,* 347-356.

McLachlan, G. J. (1992) *Discriminant analysis and statistical pattern recognition.* New York: Wiley and Sons.

Morrison, D. G. (1974). Discriminant analysis. In Ferber R. (ed.) *Handbook on marketing research*. New York: McGraw-Hill.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Rao, C. R. (1973). *Linear statistical inference and its applications.* New York: Wiley and Sons.

Rinott, Y., & Tam, M. (2003). Monotone regrouping, regression, and Simpson's paradox. *The American Statistician, 57*, 139-141.

Ripley, B. D. (1996). *Pattern recognition and neural networks.* London: Cambridge University Press.

Rosenbaum, P. R. (1995) *Observational studies.* New York: Springer-Verlag.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. New York: Chapman & Hall.

Wainer, H., & Brown, L. M. (2004). Two statistical paradoxes in the interpretation of group differences: illustrated with medical school admission and licensing data. *The American Statistician, 58*, 117-123.

Winship, C., & Morgan, L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659-707.

Appendix:

The Sherman-Morrison formula

$$(A+uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1+u'A^{-1}v} \qquad \text{(A1)}$$

is well known in various theoretical and practical statistical evaluations. It is convenient to use when the inverted matrix A^{-1} is already known, so the inversion of $A+uv'$ can be expressed via A^{-1} due to the formula (A1).

We extend this formula to the inversion of a matrix with two pairs of vectors. Consider a matrix $A+u_1v_1'+u_2v_2'$, where *A* is a square non-singular matrix of *n*-th order, and $u_1v_1'+u_2v_2'$ is a matrix of the rank 2, arranged via two outer products u_1v_1' and u_2v_2' of the vectors of *n*-th order. Suppose we need to invert such a matrix to solve a linear system:

$$(A+u_1v_1'+u_2v_2')a = b\,, \qquad \text{(A2)}$$

where *a* is a vector of unknown coefficients and *b* is a given vector. Opening the parentheses, we get an expression:

$$Aa + u_1k_1 + u_2k_2 = b\,, \qquad \text{(A3)}$$

where k_1 and k_2 are unknown parameters defined as scalar products of the vectors:

$$k_1 = (v_1'a)\,, \quad k_2 = (v_2'a)\,, \qquad \text{(A4)}$$

Solution *a* can be found from (A3) as:

$$a = A^{-1}b - k_1A^{-1}u_1 - k_2A^{-1}u_2\,. \qquad \text{(A5)}$$

Substituting the solution (A5) into the system (A2) and opening the parentheses yields a vector equation:

$$\begin{aligned} &k_1u_1 + k_2u_2 + k_1q_{11}u_1 + k_2q_{12}u_1 \\ &+k_1q_{21}u_2 + k_2q_{22}u_2 = c_1u_1 + c_2u_2 \end{aligned}\,, \qquad \text{(A6)}$$

where the following notations are used for the known constants defined by the bilinear forms:

$$\begin{aligned} q_{11} &= v_1'A^{-1}u_1, \quad q_{12} = v_1'A^{-1}u_2, \\ q_{21} &= v_2'A^{-1}u_1, \quad q_{22} = v_2'A^{-1}u_2, \\ c_1 &= v_1'A^{-1}b, \quad c_2 = v_2'A^{-1}b. \end{aligned} \qquad \text{(A7)}$$

Considering equations (A6) by the elements of vector u_1 and by the elements of vector u_2, we obtain a system with two unknown parameters k_1 and k_2:

$$\begin{cases} (1+q_{11})k_1 + q_{12}k_2 = c_1 \\ q_{21}k_1 + (1+q_{22})k_2 = c_2 \end{cases}. \qquad \text{(A8)}$$

So the solution for the parameters (A4) is:

$$\begin{aligned} k_1 &= (c_1 + q_{22}c_1 - q_{12}c_2)/\Delta, \\ k_2 &= (c_2 + q_{11}c_2 - q_{21}c_1)/\Delta, \end{aligned} \qquad \text{(A9)}$$

with the main determinant of the system:

$$\begin{aligned} \Delta &= (1+q_{11})(1+q_{22}) - q_{12}q_{21} \\ &= (1+v_1'A^{-1}u_1)(1+v_2'A^{-1}u_2)\text{-}(v_1'A^{-1}u_2)(v_2'A^{-1}u_1) \end{aligned}. \qquad \text{(A10)}$$

Using the obtained parameters (A9) in the vector *a* (A5), we get:

$$a = \left\{ A^{-1} - \frac{\begin{array}{c} A^{-1}u_1v_1'A^{-1}(1+q_{22}) + A^{-1}u_2v_2'A^{-1}(1+q_{11}) \\ -A^{-1}u_1v_2'A^{-1}q_{12} - A^{-1}u_2v_1'A^{-1}q_{21} \end{array}}{\Delta} \right\} b \qquad \text{(A11)}$$

with the constants defined in (A7).

The expression in the figure parentheses (A11) defines the inverted matrix of the system (A2). It can be easily proved by multiplying the matrix in (A2) by the matrix in (A11), that yields the uniform matrix. In a simple case when both pairs of the vectors are equal, or $u_1v_1' = u_2v_2'$, they can be denoted as $u_1v_1' = u_2v_2' = 0.5uv'$, and the expression (A12) reduces to the formula (A1). We can explicitly present the inverted matrix (A11) as follows:

$$(A+u_1v_1'+u_2v_2')^{-1}=A^{-1}-\frac{A^{-1}u_1v_1'A^{-1}+A^{-1}u_2v_2'A^{-1}}{\Delta}$$
$$+\frac{\begin{pmatrix}A^{-1}u_1v_1'A^{-1}u_2v_2'A^{-1}+A^{-1}u_2v_2'A^{-1}u_1v_1'A^{-1}\\ -A^{-1}u_1v_2'A^{-1}u_2v_1'A^{-1}-A^{-1}u_2v_1'A^{-1}u_1v_2'A^{-1}\end{pmatrix}}{\Delta}. \quad \text{(A12)}$$

For the important case of a symmetric matrix A, each of the bilinear forms (A7) can be equally presented by the transposed expression, for instance,

$$\begin{aligned} q_{11}&=v_1'A^{-1}u_1=u_1'A^{-1}v_1,\\ q_{12}&=v_1'A^{-1}u_2=u_2'A^{-1}v_1,\\ q_{21}&=v_2'A^{-1}u_1=u_1'A^{-1}v_2,\\ q_{22}&=v_2'A^{-1}u_2=u_2'A^{-1}v_2.\end{aligned} \quad \text{(A13)}$$

Using the property (A13) we simplify the numerator of the second ratio in (A12) to following:

$$\begin{aligned}&A^{-1}u_1u_2'A^{-1}v_1v_2'A^{-1}+A^{-1}u_2u_1'A^{-1}v_2v_1'A^{-1}\\ &-A^{-1}u_1u_2'A^{-1}v_2v_1'A^{-1}-A^{-1}u_2u_1'A^{-1}v_1v_2'A^{-1}\\ &=A^{-1}(u_1u_2'-u_2u_1')A^{-1}(v_1v_2'-v_2v_1')A^{-1}.\end{aligned} \quad \text{(A14)}$$

So the formula (A12) for a symmetric matrix A can be represented as:

$$(A+u_1v_1'+u_2v_2')^{-1}=A^{-1}-\frac{\begin{pmatrix}A^{-1}(u_1v_1'+u_2v_2')A^{-1}\\ -A^{-1}(u_1u_2'-u_2u_1')A^{-1}(v_1v_2'-v_2v_1')A^{-1}\end{pmatrix}}{\Delta}, \quad \text{(A15)}$$

with the determinant defined in (A10).

In a special case of the outer products of each vector by itself, when $u_1=v_1$ and $u_2=v_2$, the formula (A15) transforms into:

$$(A+u_1u_1'+u_2u_2')^{-1}=A^{-1}-\frac{\begin{pmatrix}A^{-1}(u_1u_1'+u_2u_2')A^{-1}\\ -A^{-1}(u_1u_2'-u_2u_1')A^{-1}(u_1u_2'-u_2u_1')A^{-1}\end{pmatrix}}{(1+u_1'A^{-1}u_1)(1+u_2'A^{-1}u_2)-(u_1'A^{-1}u_2)^2}. \quad \text{(A16)}$$