Vector Regression

1. We use multiple regression to interpret MDS spaces when we know additional attributes about the stimuli.
2. There are a variety of ways regression can be used. The simplest is to do “vector” regression by (multivariate) multiple regression.
   - The purpose is to find the direction (vector) through the stimulus space which is most like increasing amounts of an attribute.
   - Also appropriate for relating preference information to the MDS space. Called “Preference Mapping” (“PrefMap” for short).
3. Vector regression is (multivariate) multiple regression.
   - The vector is oriented so as to maximize the multiple correlation coefficient.
   - The multiple correlation for the vector equals the simple correlation between (a) the degree of attribute possessed by each stimulus, and (b) the “projection” of each stimulus onto the vector. (The “projections” are the “predicted values”).
   - If the “projections” coincide perfectly with the “degrees” of attributes, the correlation is 1.00.
4. Any stimuli falling along an “iso-attribute” contour project onto the same location of the vector: They are “predicted” to have equal amounts of the attribute.

Using (Multivariate) Multiple Regression for Vector Regression

1. Vector regression can be done with any (multivariate) regression program. Each attribute (or preference) variable is the response variable in a series of regression analyses which all use the MDS dimensions as the predictor variables.
2. The multiple correlation coefficient tells us how strongly the attribute (preference) relates to the stimulus space.
   - The *squared* correlation is the proportion of variance in the attribute (preference) that is related to the MDS space.
3. To plot each attribute (preference) vector, use the *raw* (yes, *raw*, <u>not</u> standardized, as it says in Schiffman, et al.) regression coefficients as the coordinates of a point in the MDS space. Draw an arrow starting at the mid-point of the MDS space (origin, if the dimensions are centered) to the point, putting the head of the arrow at the point.
   - The direction of this arrow is the direction of increasing attribute (preference).
   - The length of this arrow corresponds to the correlation. (Short arrows correspond to low correlations and weak relationships between attribute (preference) and the MDS space.
4. Note: If the response on the attribute (preference) is in terms of *decreasing* amount, then the arrow points in the direction of low attribute (preference).

Ideal Point Regression
(External Unfolding, Response Surface Regression)

1. "Ideal Point" regression is a special type of (multivariate) multiple regression.
   - The purpose is to find the point in the stimulus space which is most like an attribute. As one gets further from this point, one is at locations that are less and less like the attribute.
   - Also appropriate for relating preference information to the MDS space. This is still called "Preference Mapping" ("PrefMap" for short).
   - This is called "external unfolding" since the model involves locating an attribute or preference point in an already existing space of stimulus points. We end up with a single space with two sets of points: stimulus points (which were already there) and attribute/preference points (which are based on information "external" to the original stimulus space).
2. Point regression involves (multivariate) multiple regression.
   - The point is located so as to maximize the simple correlation between the response variable and the squared Euclidean distances from the point to each of the observation (stimulus) points.
   - The multiple correlation equals that simple correlation and it tells us how well the point fits into the pre-existing stimulus space.
3. There are "iso-attribute/preference" contours which are either circles or ellipses, depending on which of the three models described below is being used. All stimulus points on the same contour have the same degree of predicted attribute/preference.

Using (Multivariate) Multiple Regression for the Circular Ideal Point Regression Model

1. There are three ideal point regression models. They each are based on (multivariate) multiple regression where:
   - The response variable(s) is/are the attribute or preference rating variable(s). This is the same as for vector regression.
   - The predictor variables consist of the stimulus space dimensions ***plus*** additional variable(s) computed from these dimensions.

2. ***Circular Ideal Point Regression***
   - Here the iso-attribute/preference contours are circles.
   - The predictor variables are the dimensions, plus another variable whose elements are the sum of squared coordinates of each point:

$$p_i = b_0 + \sum_{a=1}^{r} b_a x_{ia} + c\left(\sum_{g=1}^{r} x^2_{ig}\right) \quad ,$$

     where p_i is the attribute/preference rating for stimulus i and x_{ia} are the coordinates of point i on dimension a in an r-dimensional space.
   - The regression coefficients are b_a (for the dimensions) and c (for the sum-of-squares variable).
   - The coordinates of the ideal point are

$$y_{ia} = \frac{b_a}{-2c}$$

A Derivation of the Circular Ideal Point Formula for the Coordinates of the Ideal point[1]

In point regression we are regressing (ignoring b_0):

$$y = b_1 x_1 + b_2 x_2 + c(x_1^2 + x_2^2)$$

which can be re-written as:

$$y = cx_1^2 + b_1 x_1 + cx_2^2 + b_2 x_2$$

which can in turn be re-written as:

$$\frac{y}{c} = x_1^2 + \frac{b_1}{c} x_1 + x_2^2 + \frac{b_2}{c} x_2$$

Completing the squares leads to:

$$\frac{y}{c} + \left(\frac{b_1}{2c}\right)^2 + \left(\frac{b_2}{2c}\right)^2 = \left[x_1^2 + \frac{b_1}{c} x_1 + \left(\frac{b_1}{2c}\right)^2\right] + \left[x_2^2 + \frac{b_2}{c} x_2 + \left(\frac{b_2}{2c}\right)^2\right]$$

which can be re-written as:

$$\frac{y}{c} + \left(\frac{b_1}{2c}\right)^2 + \left(\frac{b_2}{2c}\right)^2 = \left[x_1 + \frac{b_1}{2c}\right]^2 + \left[x_2 + \frac{b_2}{2c}\right]^2 \quad .$$

This is the equation of a circle of varying radius, since the general equation of a circle with center (g,h) and radius r is $r^2 = (x_1 - g)^2 + (x_2 - h)^2$. Thus, our equation is a circle with center and radius:

$$\left(\frac{-b_1}{2c}, \frac{-b_2}{2c}\right) \text{ and } \sqrt{\frac{y}{c} + \left(\frac{b_1}{2c}\right)^2 + \left(\frac{b_2}{2c}\right)^2} \ ,$$

with the radius varying upon the value of *y*.

---

1. By Andrea Shapiro

Using (Multivariate) Multiple Regression for the Elliptical Ideal Point Regression Models

1. *Parallel Elliptical Point Regression*

- Here the iso-attribute/preference contours are ellipses whose axes parallel the dimensions of the space.
- The predictor variables are the dimensions, plus an additional variable for each dimension which contains the squares of the coordinates. That is:

$$p_i = b_0 + \sum_{a=1}^{r} b_a x_{ia} + \sum_{g=1}^{r} c_g x^2_{ig}.$$

- The coordinates of the ideal point are

$$y_{ia} = \frac{b_a}{-2c_a}$$

2. *Non-Parallel Elliptical Point Regression*

- Here the iso-attribute/preference contours are ellipses whose axes do not parallel the dimensions of the space.
- The predictor variables are the dimensions, plus an additional variable for each unique pair of dimensions which contains the products of the coordinates. That is:

$$p_i = b_0 + \sum_{a=1}^{r} b_a x_{ia} + \sum_{g=1}^{r} \sum_{h=1}^{r} c_{gh} x_{ig} x_{ih} \text{, for } g \geq h.$$

- The coordinates of the ideal point are (in matrix algebra):

$$y_i = -\frac{1}{2} b C^{-1}$$

A Derivation of the Parallel Elliptical Ideal Point Formula for the Coordinates of the Ideal point

In point regression we are regressing (ignoring b_0):

$$y = b_1 x_1 + b_2 x_2 + c_1 x_1^2 + c_2 x_2^2$$

which can be re-written as:

$$\frac{y}{c_1 c_2} = \frac{\left(x_1^2 + \frac{b_1}{c} x_1\right)}{c_2} + \frac{\left(x_2^2 + \frac{b_2}{c} x_2\right)}{c_1}$$

Completing the squares leads to:

$$\frac{y}{c_1 c_2} + \left(\frac{b_1}{2c_1}\right)^2 + \left(\frac{b_2}{2c_2}\right)^2 = \frac{\left[x_1 + \frac{b_1}{2c_1}\right]^2}{c_2} + \frac{\left[x_2 + \frac{b_2}{2c_2}\right]^2}{c_1} \ .$$

Since the general equation of an ellipse with axes parallel to the dimensions, center (g,h) and radius r is $r^2 = \frac{(x_1 - g)^2}{a^2} + \frac{(x_2 - h)^2}{b^2}$, our equation is an ellipse with center and radius:

$$\left(\frac{-b_1}{2c_1}, \frac{-b_2}{2c_2}\right) \text{ and } \sqrt{\frac{y}{c_1 c_2} + \left(\frac{b_1}{2c_1}\right)^2 + \left(\frac{b_2}{2c_2}\right)^2} \ ,$$

with the radius varying upon the value of *y.*

Canonical Regression Analysis

1. Canonical analysis is appropriate for the situation in which there are two sets of multiple variables.
2. Canonical analysis finds linear combinations of one set that are maximally correlated with linear combinations of the other set.
3. The **scores** (not the coefficients) on the canonical combinations within a set of variables are orthogonal.
4. Multiple regression is a special case of canonical analysis in which one "set" of variables consists of only one variable.
5. Canonical analysis differs from multiple regression in that neither set of variables is "response/dependent" nor "predictor/independent".
6. Both sets of variables are viewed equivalently – both take on the same role in the analysis – both are simply "a set of variables that are linearly related to the other set of variables".
7. Canonical analysis is a very good "meta-model": i.e., a model that includes many other models as special cases.
8. Canonical analysis is not generally appropriate for data analysis because it is seldom that researchers have empirical situations that give rise to two sets of variables, neither of which is dependant on the other.
9. Because of this, I would no longer pursue the developments given in chapter 12 of SRY.

Principal Components Regression

1. As mentioned above, canonical analysis is not generally appropriate for data analysis because researchers usually have empirical situations that give rise to one set of variables that is dependant on the another.
2. If the research realizes that canonical analysis is not appropriate, the most common substitute analysis is "principal components regression".
3. This is, simply,
   A) Principal components of the predictors, followed by
   B) Multivariate multiple regression using the first "few" principal components as the predictors.
4. We can form a low dimensional biplot of the predictor space (by forming a scatterplot from the first 2 or 3 principal component scores and displaying vectors from the predictor variable coefficients).
5. We can then form a "triplot" by displaying the response variables in this space as an additional set of vectors by using the regression coefficients.
6. What we have is a picture that shows the maximum variation of the predictors, and how the two sets of variables project into this space.
7. The problem with this analysis is that it does not clearly optimize any single index. Rather, each step optimizes a separate index.
8. The appropriate analysis is redundancy analysis, which amounts to
   A) Multivariate multiple regression, followed by
   B) Principal components analysis of the predicted values.
9. This analysis clearly optimizes a specific index, as shown on the next few slides.
10. We can still form a triplot.

Redundancy Analysis

1. Redundancy analysis is also appropriate for the situation in which there are two sets of multiple variables.
2. Unlike canonical analysis, redundancy analysis finds linear combinations of one set that are maximally correlated with the other set of variables (for which no linear combination is obtained).
3. It is still the case that multiple regression is a special case of canonical analysis in which one “set” of variables only contains one variable.
4. It is also the case that redundancy analysis is a special case of canonical analysis (one of the sets of linear combinations is restricted to be the identity transformation).
5. Redundancy analysis is like multiple regression: one set of variables is “response/dependent” and the other is “predictor/independent”.
6. The variable for which no linear combination are obtained are viewed as the response to the linear combinations of the other, predictor, set.
7. Redundancy analysis is often good for data analysis involving two sets of variables since researchers often have empirical situations that give rise to two sets of variables, one of which is dependant on the other.
8. Because of this, I would use redundancy analysis in place of the developments given in chapter 12 of SRY.
9. Point versions of redundancy analysis are possible, and can be computed with PROC TRANSREG.
10. Redundancy analysis is the same as principal components of the predicted response values from multivariate multiple regression.
11. If the two sets of variables are the same, redundancy analysis becomes principal components analysis.

Redundancy Analysis Equations

1. The basic redundancy analysis equation is:

$$X = YAB + E,$$

where:

X is an (**n** x **p**) matrix of **p** centered response variables with data from n cases;

Y is an (**n** x **q**) matrix of **q** centered predictor variables with data from the same n cases;

A is a (**q** x **r**) matrix of coefficients of the **q** predictor variables on **r** ≤ **q** redundancy variables, where A is column orthogonal (i.e., $A'A = D$ is diagonal);

B is a (**r** x **p**) matrix of coefficients of the **r** ≤ **p** redundancy variables on the **p** response variables, where B is row orthogonal (i.e., $BB' = D$ is diagonal);

E is an (**n** x **p**) matrix of least squares residuals from fit of the model *YAB* to the response data *X*.

2. Note that $X \approx \tilde{X} = YC$ where $C = AB$. That is, the set of response variables X are least squares fit by the predicted values $\tilde{X}$ which are themselves linear combinations C of the predictors Y. Note that the linear combinations have structure $C = AB$.

3. We can define $Z = YA$ where $Z'Z = \Delta$ is a diagonal matrix with diagonal elements δ_i such that $\mathrm{Max}(\delta_1)$ and $\mathrm{Max}\langle\delta_i|\delta_{1\dots i-1}\rangle, i = 1\dots r$.

   Then $\tilde{X} = ZB$ is the decomposition of $\tilde{X}$ into

   Z, the matrix of scores on the redundancy variables, and

   B, the matrix of coefficients of the redundancy variables.