Polynomial Regression and Response Surface Methodology

Jeffrey R. Edwards

University of North Carolina

In C. Ostroff & T. A. Judge (Eds.), *Perspectives on organizational fit* (pp. 361-372). San Francisco: Jossey-Bass.

Person-environment (P-E) fit research often relies on methods that collapse person and environment measures into a single score intended to represent P-E fit. Typically, these methods involve computing the difference between person and environment measures or the similarity between profiles of measures that describe person and environment on multiple dimensions. These methods suffer from numerous methodological problems, as documented elsewhere (Cronbach, 1958, 1992; Edwards, 1994; Johns, 1981). Problems with difference scores and profile similarity indices are avoided by polynomial regression (Edwards, 1994, 2002; Edwards & Parry, 1993), which uses separate measures of the person and environment and examines their joint relationships with causes and consequences of P-E fit. Polynomial regression is based on the premise that person and environment measures represent distinct constructs and the assumptions embedded in difference scores and profile similarity indices represent hypotheses that should be tested empirically.

This section discusses polynomial regression and its relevance to P-E fit research. The section has three objectives. First, it shows how polynomial regression can be viewed as a generalization of difference scores and profile similarity indices. Second, it explains how results from polynomial regression analyses can be understood using response surface methodology. Third, it emphasizes that polynomial regression and response surface methodology can facilitate theory development in P-E fit research. This overview focuses on fit as a cause of outcomes, and procedures for treating fit as an outcome are briefly discussed at the end of the summary.

Polynomial Regression as a Generalization of Difference Scores and Profile Similarity Indices

The basics of polynomial regression can be understood by contrasting it with difference scores and profile similarity indices. To illustrate, consider Figure 1a, which depicts a positive relationship between an algebraic difference score and an outcome. This relationship can be represented by the following regression equation:

$$Z = b_0 + b_1(X - Y) + e \quad (1)$$

where X is the environment, Y is the person, Z is the outcome, and e is a random disturbance term. The positive relationship in Figure 1a corresponds to a positive value for b_1 in Equation 1. The connection between Equation 1 and polynomial regression can be seen by expanding Equation 1, which yields:

$$Z = b_0 + b_1X - b_1Y + e. \quad (2)$$

Equation 2 shows that using an algebraic difference as a predictor is equivalent to using the components of the difference as predictors and constraining their coefficients to be equal in magnitude and opposite in sign. The relationship of X and Y with Z indicated by Equation 2 is illustrated by the three-dimensional surface in Figure 2a. The constraint imposed by Equation 2 can be empirically tested using the following equation:

$$Z = b_0 + b_1X + b_2Y + e. \quad (3)$$

Equation 3 is a linear polynomial regression equation in which the relationships of X and Y with Z can differ in sign and magnitude. Results from Equation 3 can be used to determine whether $b_1 = -b_2$, as indicated by Equation 2 , and whether b_1 is positive and b_2 is negative, as implied by Figure 2a.

Figure 1b shows an inverted-V relationship, such that the outcome is maximized when X and Y are equal. The relationship in Figure 1b is captured by the following regression equation:

$$Z = b_0 + b_1|X - Y| + e. \quad (4)$$

Equation 4 uses the absolute value of the difference between X and Y as a predictor of Z. A negative value for b_1 would correspond to the inverted-V relationship in Figure 1b. As written, Equation 4 cannot be algebraically expanded, because an absolute difference is a logical rather than an algebraic transformation. This dilemma is overcome by replacing Equation 4 with the following equation (Edwards, 1994):

$$Z = b_0 + b_1(1 - 2W)(X - Y) + e. \quad (5)$$

In Equation 5, W is a dummy variable that equals 0 when X – Y is positive, equals 1 when X – Y is negative, and is randomly set to 0 or 1 when X = Y. As a result, when X – Y is positive, 1 – 2W equals 1, and the compound term (1 – 2W)(X – Y) reduces to (X – Y). In contrast, when X – Y is negative, 1 – 2W equals –1, and the term (1 – 2W)(X – Y) becomes –(X – Y). Hence, when X – Y is positive, its sign is unaltered, and when X – Y is negative, its sign is reversed, producing the same result as the absolute value transformation. When X = Y, (1 – 2W)(X – Y) equals zero regardless of the value of W. Expanding Equation 5 yields:

$$Z = b_0 + b_1X - b_1Y - 2b_1WX + 2b_1WY + e. \qquad (6)$$

Equation 6 shows that using an absolute difference as a predictor effectively constrains the coefficients on X and Y to be equal in magnitude but opposite in sign, constrains the coefficients on WX and WY to be equal in magnitude but opposite in sign, and constrains the coefficient on WY to be twice as large as the coefficient on X. Equation 6 also indicates that the coefficient on W is constrained to zero, given that it is excluded from the equation. Figure 2b shows a three-dimensional surface that corresponds to Equation 6. The constraints imposed by Equation 6 can be tested with the following equation:

$$Z = b_0 + b_1X + b_2Y + b_3W + b_4WX + b_5WY + e. \qquad (7)$$

Equation 7 is a piecewise polynomial regression equation. Coefficient estimates this equation can be used to assess the constraints imposed by Equation 6 by testing whether: (a) $b_1 = -b_2$; (b) $b_4 = -b_5$; (c) $b_5 = 2b_1$; and (d) $b_3 = 0$. The direction of the inverted-V relationship in Figure 2b further stipulates that b_1 and b_5 are positive and that b_2 and b_4 are negative.

Figure 1c shows an inverted-U relationship. Like Figure 1b, Figure 1c indicates that the outcome is maximized when X and Y are equal. However, as the difference between X and Y increases, the outcome decreases linearly in Figure 1b, as opposed to the curvilinear decrease in Figure 1c. The relationships shown in Figures 1b and 1c are usually treated the same in P-E fit research. This relationship in Figure 1c corresponds to the following regression equation:

$$Z = b_0 + b_1(X - Y)^2 + e. \quad (8)$$

Equation 8 uses the squared difference between X and Y as a predictor of Z. For the relationship in Figure 1c, the sign of b_1 in Equation 8 would be negative. Expanding Equation 8 yields:

$$Z = b_0 + b_1X^2 - 2b_1XY + b_1Y^2 + e \quad (9)$$

Equation 9 indicates that using a squared difference as a predictor constrains the coefficients on X^2 and Y^2 to be equal and the coefficient on XY to be twice as large in magnitude and opposite in sign of the coefficient on X^2 or Y^2. Equation 9 also implicitly constrains the coefficients on X and Y to be zero, given that both of these variables are excluded from Equation 9. A three-dimensional surface corresponding to Equation 9 is shown in Figure 2c, and the four constraints imposed by Equation 9 can be tested with the following equation:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e. \quad (10)$$

Equation 10 is a quadratic polynomial regression equation. Coefficient estimates from Equation 10 can be used to evaluate the constraints imposed by Equation 9 by testing whether: (a) $b_1 = 0$; (b) $b_2 = 0$; (c) $b_3 = b_5$; and (d) $b_4 = -2b_3$. The direction of the inverted-U relationship in Figure 2b further indicates that b_3 and b_5 are negative and b_4 is positive.

The logic used to translate difference scores into polynomial regression equations can be applied to profile similarity indices (Cronbach & Gleser, 1953; Edwards, 1993, 1994). For example, profile similarity indices that represent sums of algebraic, absolute, or squared differences can be written by extending Equations 2, 6, and 9, respectively, to include multiple pairs of X and Y measures in which each pair represents a dimension on which the profiles are compared. The constraints imposed by these indices can be tested using extended versions of Equations 3, 7, and 10. Euclidean distance and profile correlation indices, which are commonly used in P-E fit research, cannot be algebraically expanded, but the conceptual principles they are intended to capture can be examined using unconstrained regression equations for the sums of absolute or squared differences (Edwards, 1993).

Applying Response Surface Methodology to Polynomial Regression Analysis

When polynomial regression yields coefficients that satisfy the constraints associated with Equations 2, 6, and 9, results are easily interpreted because they conform to the idealized surfaces shown in Figure 2. However, these constraints are usually rejected, which complicates the interpretation of results. Furthermore, the surfaces in Figure 2 comprise a narrow subset of hypotheses that could be developed regarding the joint effects of the person and environment on outcomes. For example, outcomes produced by P-E misfit may differ depending on whether the environment is greater than or less than the person (Edwards, Caplan, & Harrison, 1998; Naylor, Pritchard, & Ilgen, 1980; Rice, McFarlin, Hunt, & Near, 1985). In addition, the effects of P-E fit may depend on whether the person and environment are both low or high in an absolute sense (Edwards & Rothbard, 1999). Complexities such as these are important from a theoretical standpoint but are not captured by the surfaces in Figure 2.

The foregoing issues can be addressed using response surface methodology (Edwards & Parry, 1993), which allows researchers to rigorously analyze three-dimensional surfaces relating the person and environment to outcomes. Response surface methodology facilitates substantive interpretation when constraints imposed by difference scores are rejected, as is usually the case. Perhaps more importantly, response surface methodology allows P-E fit researchers to develop and test hypotheses that go far beyond the simplified surfaces shown in Figure 2.

Response surface methodology involves analyzing features of surfaces that correspond to polynomial regression equations. The quadratic equation in Equation 10 captures a wide range of hypotheses relevant to P-E fit research and is therefore the focus of the present discussion. A quadratic equation reflects one of three types of surfaces: (a) *concave*, which is dome-shaped; (b) *convex*, which is bowl-shaped; and (c) *saddle,* which is shaped like a saddle. For each surface, response surface methodology involves the analysis of three basic features.

The first feature is the *stationary point*, which is the point at which the surface is flat. For a concave surface, the stationary point is the overall maximum of the surface. For a convex surface, the stationary point is the overall minimum of the surface. For a saddle surface, the stationary point is the intersection of the lines along which the upward and downward curvatures of the surface are greatest. The location of the stationary point can be computed by inserting the estimated coefficient values from a quadratic regression equation into the following formulas:

$$X_0 = \frac{b_2 b_4 - 2b_1 b_5}{4b_3 b_5 - b_4^2} \quad (11)$$

$$Y_0 = \frac{b_1 b_4 - 2b_2 b_3}{4b_3 b_5 - b_4^2} \quad (12)$$

where X_0 and Y_0 are the coordinates of the stationary points in the X,Y plane.

The second feature involves the *principal axes,* which describe the orientation of the surface in the X,Y plane. The principal axes run perpendicular to one another and intersect at the stationary point. For a concave surface, the first principal axis is the line of minimum downward curvature, and the second principal axis is the line of maximum downward curvature. For a convex surface, the first principal axis is the line of maximum upward curvature, and the second principal axis is the line of minimum upward curvature. Finally, for a saddle surface, the first principal axis is the line of maximum upward curvature, and the second principal axis is the line of maximum downward curvature.

The principal axes can be written as equations that describe lines in the X,Y plane. An equation for the first principal axis is:

$$Y = p_{10} + p_{11}X. \quad (13)$$

The slope of the first principal axis (i.e., p_{11}) is computed as follows:

$$p_{11} = \frac{b_5 - b_3 + \sqrt{(b_3 - b_5)^2 + b_4^2}}{b_4}. \quad (14)$$

The intercept of the first principal axis (i.e., p_{10}) can be computed as follows:

$$p_{10} = Y_0 - p_{11}X_0. \tag{15}$$

Likewise, an equation for the second principal axis is:

$$Y = p_{20} + p_{21}X. \tag{16}$$

The slope of the second principal axis (i.e., p_{21}) is computed using the following formula:

$$p_{21} = \frac{b_5 - b_3 - \sqrt{(b_3 - b_5)^2 + b_4^2}}{b_4} \tag{17}$$

The intercept of the second principal axis (i.e., p_{20}) is computed as follows:

$$p_{20} = Y_0 - p_{21}X_0. \tag{18}$$

The third feature entails the shape of the surface along relevant lines in the X,Y plane, which can be computed by substituting the equation for the line into Equation 10. To illustrate, the Y = X line is meaningful to P-E fit research because it represents values where the person and environment are equal. For the surface in Figure 2c, this line runs diagonally across the floor of each graph from the near corner to the far corner. Substituting Y = X into Equation 10 yields:

$$\begin{aligned} Z &= b_0 + b_1X + b_2X + b_3X^2 + b_4X^2 + b_5X^2 + e \\ &= b_0 + (b_1 + b_2)X + (b_3 + b_4 + b_5)X^2 + e. \end{aligned} \tag{19}$$

As Equation 19 shows, the curvature of the surface along the Y = X line is represented by the sum $b_3 + b_4 + b_5$, and the slope of the surface at the point X = 0 (and Y = 0, given that Y = X) is $b_1 + b_2$. If these sums equal zero, then the surface is flat along the Y = X line, consistent with the surface in Figure 2c.

Another line of interest in P-E fit research is the Y = –X line. When X and Y measures are centered at the midpoint of their scales, as recommended for polynomial regression analysis (Edwards, 1994; Edwards & Parry, 1993), the Y = –X line runs diagonally across the X,Y plane and represents varying degrees of P-E misfit. In Figure 2, the Y = -X line extends from the left corner to the right corner of the floor of each graph. The shape of the surface along this line represents the effect of P-E misfit. Substituting Y = –X into Equation 10 yields:

$$Z = b_0 + b_1X - b_2X + b_3X^2 - b_4X^2 + b_5X^2 + e.$$

$$= b_0 + (b_1 - b_2)X + (b_3 - b_4 + b_5)X^2 + e. \qquad (20)$$

The curvature of the surface along the Y = –X line equals $b_3 - b_4 + b_5$, and the slope of the surface when X = 0 (and Y = 0, given than Y = –X) equals $b_1 - b_2$. For the surface in Figure 2c, $b_3 - b_4 + b_5$ is negative and $b_1 - b_2$ equals zero and, meaning that the surface has a downward curvature along the Y = –X line and is flat at X = 0, Y = 0.

The shape of the surface along other lines can be determined in a similar manner. For instance, the shape of the surface along the first principal axis is found by substituting Equation 13 into Equation 10, which yields:

$$Z = b_0 + b_1X + b_2(p_{10} + p_{11}X) + b_3X^2 + b_4X(p_{10} + p_{11}X) + b_5(p_{10} + p_{11}X)^2 + e$$

$$= b_0 + b_2p_{10} + b_5p_{10}^2 + (b_1 + b_2p_{11} + b_4p_{10} + 2b_5p_{10}p_{11})X + (b_3 + b_4p_{11} + b_5p_{11}^2)X^2 + e. \qquad (21)$$

Likewise, the shape of the surface along the second principal axis is:

$$Z = b_0 + b_1X + b_2(p_{20} + p_{21}X) + b_3X^2 + b_4X(p_{20} + p_{21}X) + b_5(p_{10} + p_{11}X)^2 + e$$

$$= b_0 + b_2p_{20} + b_5p_{20}^2 + (b_1 + b_2p_{21} + b_4p_{20} + 2b_5p_{20}p_{21})X + (b_3 + b_4p_{21} + b_5p_{21}^2)X^2 + e. \qquad (22)$$

This procedure can be extended to other lines of theoretical interest.

Empirical Example

Figure 3 depicts a response surface based on data from 358 job seekers who reported the actual variety, desired variety, and satisfaction associated with jobs for which they had recently interviewed. The estimated polynomial regression was:

$$Z = 5.628 + 0.314X - 0.118Y - .145X^2 + .299XY - .102Y^2 + e. \qquad (23)$$

The R^2 for the equation was .162, and coefficients for all variables except Y and Y^2 were statistically significant at $p < .05$. The corresponding surface is shown in Figure 3. The stationary point was located at X = –0.951, Y = –1.973 The first principal axis had an intercept and slope of –0.875 and 1.154, respectively, and is represented by the solid line crossing the X,Y plane. The second principal axis had an intercept and slope of –2.797 and –0.866 and is depicted by the heavy dashed line in the X,Y plane. Along the Y = X line, the surface had a curvature of 0.052 and a slope of 0.196 at the point X = 0, Y = 0, indicating that satisfaction increased at an increasing rate. Along the Y = –X line, the surface had a curvature of –0.546 and a slope of 0.432 at the point X = 0, Y = 0. These results indicate that satisfaction increased as actual variety increased toward desired variety, continued to increase at the point where actual and desired variety were equal (i.e., X = 0, Y = 0), and began to decrease when actual variety exceeded desired variety by about 0.5 units, as indicated by the point where the Y = –X line crossed the first principal axis.

Comparing the surface in Figures 3 to Figure 2c highlights two key findings revealed by polynomial regression and response surface methodology that would have been missed by using a squared difference score. First, Figure 3 shows that, along the line of P-E fit, satisfaction is higher when actual and desired variety are both high rather than low, whereas in Figure 2c, satisfaction is forced to remain constant along the P-E fit line. The increase in satisfaction as actual and desired variety increase makes sense from a conceptual standpoint, given that jobs with higher variety often bring additional rewards such as autonomy and challenge, and people who value variety are also likely to value these rewards. Second, Figure 3 shows that, along the line of P-E misfit, satisfaction is greatest when actual variety exceeds desired variety, while in Figure 2c, satisfaction is constrained to reach its maximum when actual and desired variety are equal. Again, the fact that satisfaction is greatest when actual variety exceeds desired variety makes sense, because a moderate excess of variety brings opportunities for challenge and self-development, which can increase overall satisfaction with the job.

The Theoretical Value of Polynomial Regression and Response Surface Methodology

The incremental insights yielded by polynomial regression and response surface methodology are important for interpreting results as well as for developing theory. By conceptualizing the effects of P-E fit as a three-dimensional surface rather than a two-dimensional function, an enormous range of hypotheses can be pursued. These hypotheses can address asymmetries in the effects of P-E misfit, variation in outcomes along the line of P-E fit, surface rotations indicating that the optimal combination of P and E depends on whether both are high or low, and so forth (Edwards, 1996; Edwards & Rothbard, 1999). Traces of hypotheses such as these are evident in the P-E fit literature (Edwards et al., 1998; Rice et al., 1985), but efforts to develop and test them have been hampered by the use of difference scores. Polynomial regression and response surface methodology encourage P-E fit researchers to adopt an expanded conceptualization of P-E fit that opens new avenues for theory and research.

Summary

Polynomial regression has numerous advantages over difference scores and profile similarity indices. From a methodological standpoint, polynomial regression avoids problems with difference scores and profile similarity indices that have plagued P-E fit research for decades. From a conceptual standpoint, polynomial regression can be coupled with response surface methodology to gain new insights from data analysis and pursue hypotheses that have been neglected in P-E fit research. The use of polynomial regression and response surface methodology has gained momentum in studies of P-E fit, and its continued use holds promise for the advancement of P-E fit theory and research.

References

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353-379). Stanford, CA: Stanford University Press.

Cronbach, L. J. (1992). Four *Psychological Bulletin* articles in perspective. *Psychological Bulletin, 112,* 389-392.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing the similarity between profiles. *Psychological Bulletin, 50,* 456-473.

Edwards, J. R. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. *Personnel Psychology*, *46*, 641-665.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58,* 51-100 (erratum, *58,* 323-325).

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64,* 307-324.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.

Edwards, J. R., Caplan, R. D., & Harrison, R. V. (1998). Person-environment fit theory: Conceptual foundations, empirical evidence, and directions for future research. In C. L. Cooper (Ed.), *Theories of organizational stress* (pp. 28-67). Oxford: Oxford University Press.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management*

Journal, 36, 1577-1613.

Edwards, J. R., & Rothbard, N. P. (1999). Work and family stress and well-being: An examination of person-environment fit in the work and family domains. *Organizational Behavior and Human Decision Processes, 77,* 85-129.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance, 27,* 443-463.

Naylor, J. C., Pritchard, R. D., & Ilgen, D. R. (1980). *A theory of behavior in organizations*. New York: Academic Press.

Rice, R. W., McFarlin, D. B., Hunt, R. G., & Near, J. P. (1985). Organizational work and the perceived quality of life: Toward a conceptual model. *Academy of Management Review, 10,* 296-310.

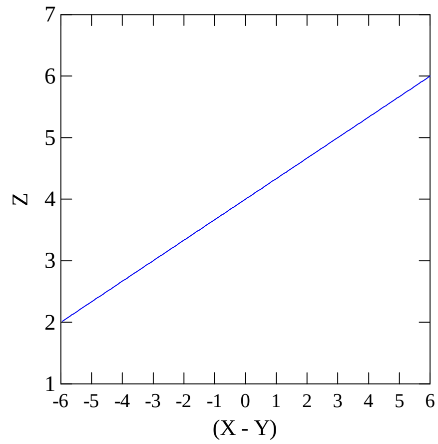
Figure Captions

Figure 1. Two-dimensional difference score functions.

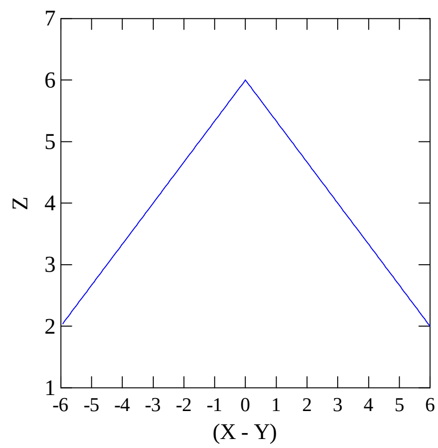
Figure 2. Three-dimensional difference score surfaces.

Figure 3. Response surface analysis relating actual and desired variety to satisfaction.

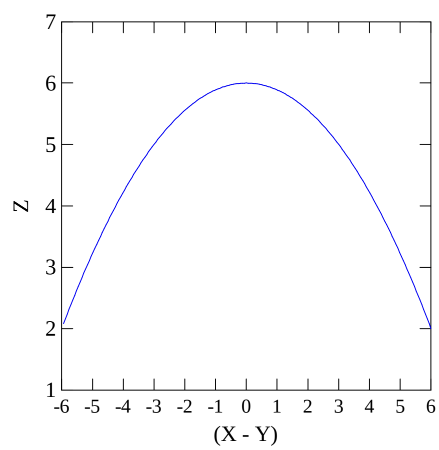
a. Two-Dimensional Algebraic Difference Function



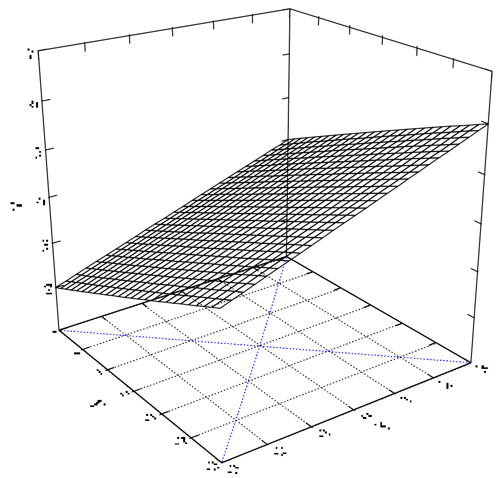
b. Two-Dimensional Absolute Difference Function



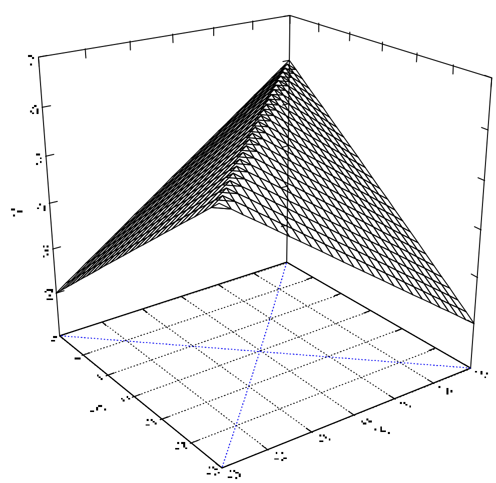
c. Two-Dimensional Squared Difference Function



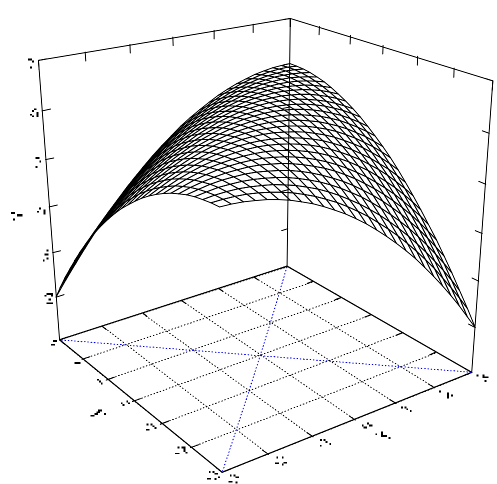
a. Three-Dimensional Algebraic Difference Function

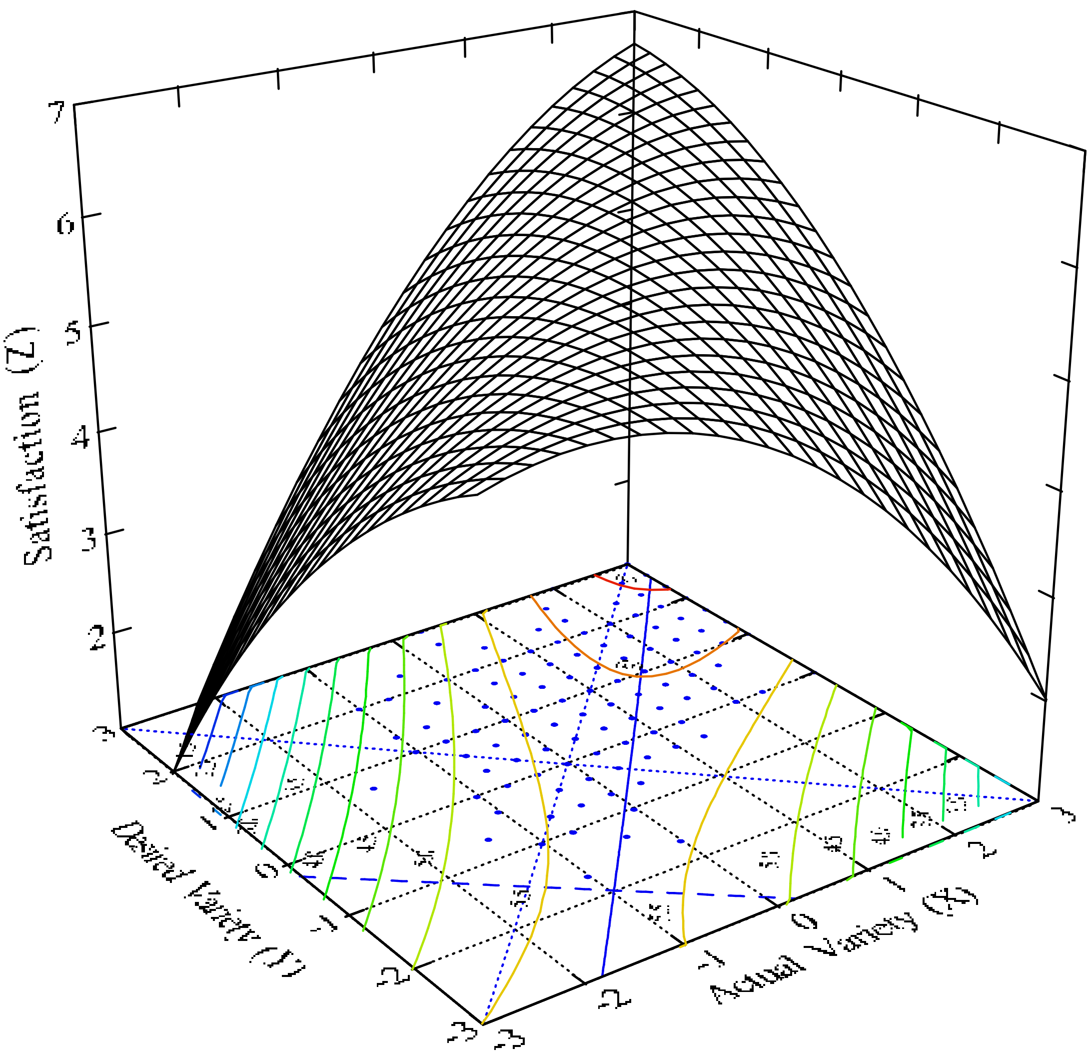


b. Three-Dimensional Absolute Difference Function



c. Three-Dimensional Squared Difference Function




Satisfaction (Z)
Desired Variety (Y)
Actual Variety (X)