

CHAPTER 8

POLYNOMIAL REGRESSION

- Used when data indicates a curvilinear response function.
- Extrapolation beyond the range of data is not recommended.
- Example. Second order model in one independent variable:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \epsilon_i$$

with $x_i = X_i - \bar{X}$ so as to reduce multicollinearity and avoid computational difficulties.

- Example. Second order model with two independent variables:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \epsilon_i$$

with $x_{i1} = X_{i1} - \bar{X}_1$ and $x_{i2} = X_{i2} - \bar{X}_2$.

The coefficients β_1 and β_2 are *linear effect* coefficients,

β_{11} and β_{22} are *quadratic effect* coefficients, and

β_{12} is the *interaction effect* coefficient for the cross product term.

Response surface is three-dimensional. See examples and figures in text.

- For an example of a third order model see text. Models of order higher than the third are seldom used. Often one fits a second or third order model and then explores whether a lower order is adequate.
- Hierarchical approach to fitting polynomial model: fit the higher order model first, then explore whether a lower-order model is adequate.

QUALITATIVE PREDICTORS

Example. If gender is a second predictor variable in a model, one may use indicator variable

$$X_2 = \begin{cases} 1 & \text{if person is female} \\ 0 & \text{if person is male} \end{cases}$$

If a qualitative predictor variable has k categories, then $k - 1$ indicator variables are needed.

Example. Suppose a second predictor variable has categories urban, suburban, and rural.

The one may use $X_2 = \begin{cases} 1 & \text{if urban} \\ 0 & \text{otherwise} \end{cases}$

$$X_3 = \begin{cases} 1 & \text{if suburban} \\ 0 & \text{otherwise} \end{cases}$$

If a model consists of only qualitative predictor variables, it is an *analysis of variance* model.