Linear Statistical Modeling Methods with SAS

Polynomial Regression Models and Qualitative Predictors

Xuemao Zhang
East Stroudsburg University

March 13, 2024

Outline

- General MLR models
- Transformations
- Polynomial Regression Models
- Qualitative Predictors

General MLR models

We assume the response Y is related to p **predictors** $Z_1, Z_2, \ldots, Z_p$ through k **regressors** $X_1(Z_1, Z_2, \ldots, Z_p), X_2(Z_1, Z_2, \ldots, Z_p), \ldots, X_k(Z_1, Z_2, \ldots, Z_p)$ and the linear relation

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1(Z_1, Z_2, \ldots, Z_p) \\ &\quad + \beta_2 X_2(Z_1, Z_2, \ldots, Z_p) \\ &\quad + \ldots + \beta_k X_k(Z_1, Z_2, \ldots, Z_p) + \epsilon, \end{aligned}$$

where ϵ is a random error with mean 0. Usually it is assumed that $\epsilon \sim N(0, \sigma^2)$.

Again,we write the general MLR model as

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon}.$$

The deterministic part of the model

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

describes average value of Y for any fixed values of $x_1, x_2, \ldots, x_k$.

General MLR models

Here are some examples:

$$
\begin{aligned}
(\text{MLR with one predictor })Y &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_1^2 + \epsilon, \\
&\quad (p = 1,\ k = 2,\ X_1 = Z_1,\ X_2 = Z_1^2) \\
(\text{Full quadratic MLR model})Y &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1^2 \\
&\quad + \beta_4 Z_1 Z_2 + \beta_5 Z_2^2 + \epsilon, \\
&\quad (p = 2,\ k = 5,\ X_1 = Z_1,\ X_2 = Z_2, \\
&\quad X_3 = Z_1^2,\ X_4 = Z_1 Z_2,\ X_5 = Z_2^2) \\
(\text{MLR with two predictors })Y &= \beta_0 + \beta_1 \log(Z_2) + \beta_3 \sqrt{Z_1 Z_2} + \epsilon. \\
&\quad (p = 2,\ k = 2,\ X_1 = \log(Z_2),\ X_2 = \sqrt{Z_1 Z_2})
\end{aligned}
$$

Remark. A **linear model** is defined as a model that is linear in the parameters, i.e., linear in the coefficients, the β's. So all the above models are linear models.

Transformations

- If there is a non-linear relationship between the response Y and the independent variable X, a SLR model between Y and X cannot be justified. We consider the use of transformations of one or both of the original variables before carrying out the regression analysis. Simple transformations of either the response variable Y or the predictor variable X, or of both, are often sufficient to make the **simple linear regression model** appropriate for the transformed data.
- **Example.** Among the factors affecting the performance of automobile engines is the air-fuel ratio, the mixture of air and fuel fed into the engine. A normalized measure of this ratio is the equivalence ratio, defined as the quotient (stoichiometric air-fuel ratio)/(actual air-fuel ratio). The higher the equivalence ratio, the richer the air-fuel mixture.
- On the next slide is a plot of fuel consumption, in μ_g/J, versus equivalence ratio for a set of experimental data(fuel). How would you summarize the relation?

Transformations

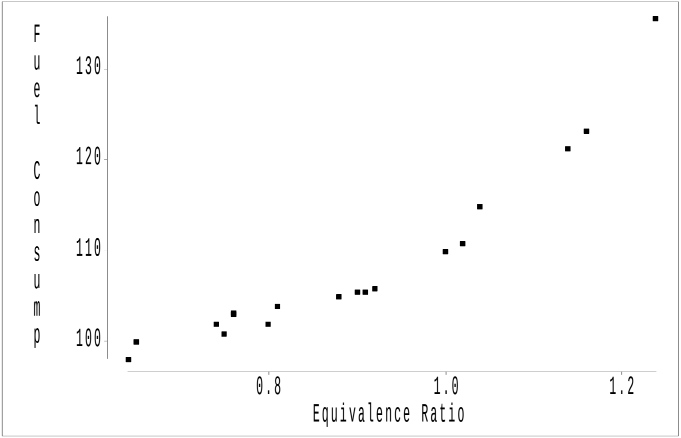
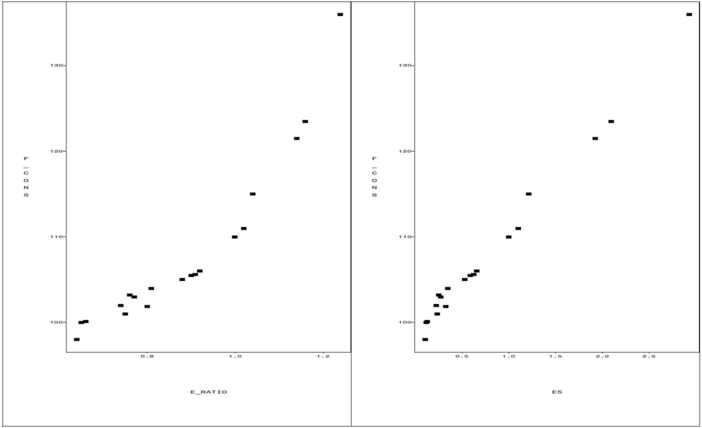


Figure 1: *Fuel consumption versus equivalence ratio for a set of experimental data.*

Transformations

Here is a plot of fuel consumption versus equivalence ratio (left) and versus the fifth power of equivalence ratio (right).



Transformations

The plot on the right seems to show a nearly linear association. In terms of the SLR model, we have

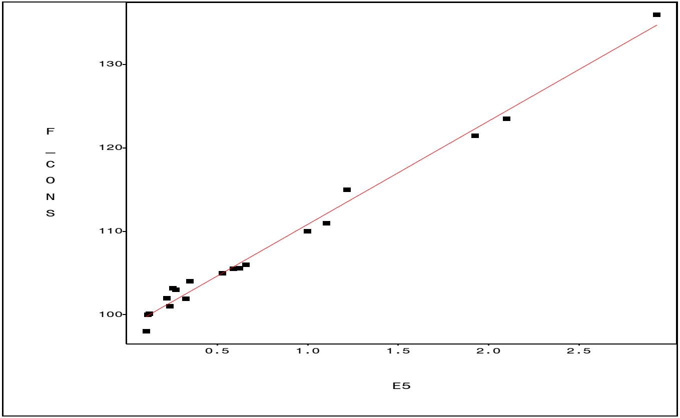
$$\begin{aligned} Y &= \beta_0 + \beta_1 X(Z) + \epsilon \\ &= \beta_0 + \beta_1 Z^5 + \epsilon \end{aligned}$$

Where the response Y is fuel consumption, the predictor Z is equivalence ratio, and the regressor $X(Z)$ is Z^5, the fifth power of equivalence ratio.

The relevant SAS output for the regression looks like this:

Transformations

Model Equation					
F_CONS	=	98.4483	+	12.3764	E5

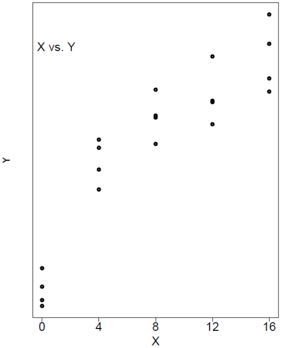


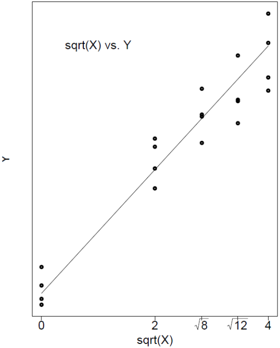
Transformations - Nonlinear Relation Only

Situation 1: nonlinear regression function with con stant error variances:

- $E(Y)$ doesn't appear to be a linear function of X, that is, the points do not seem to lie on a line.
- However, the spread of the Y's at each level of X appears to be constant, however.

We consider transformation of X only. Do not transform Y because this will disturb the spread of the Y's at each level X.



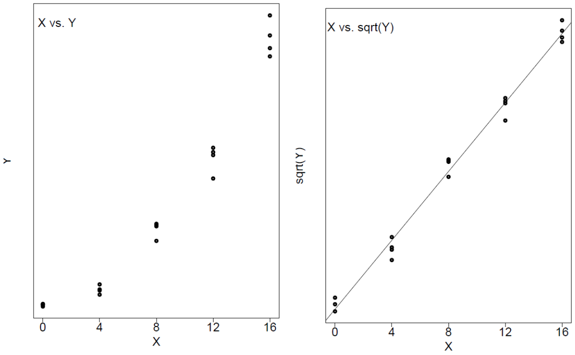


Transformations - Nonlinear Relation and Non-constant Variances

Situation 2: nonlinear regression function with non-constant error variances:

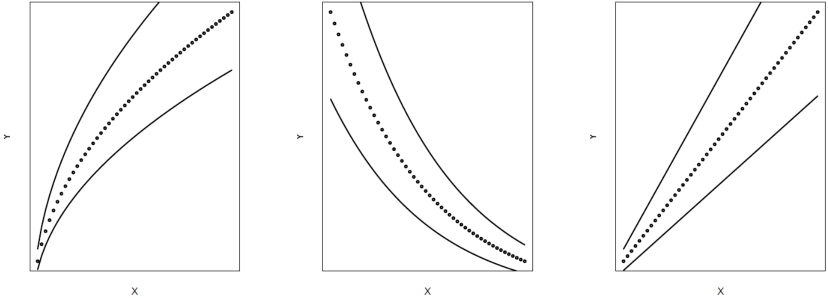
- $E(Y)$ is not a linear function of X, that is, the points do not seem to lie on a line.
- The spread of the Y's at each level of X is not a constant.

We consider transformation of Y or maybe both X and Y and hope that both problems are fixed.



Transformations - Prototypes for Transforming Y

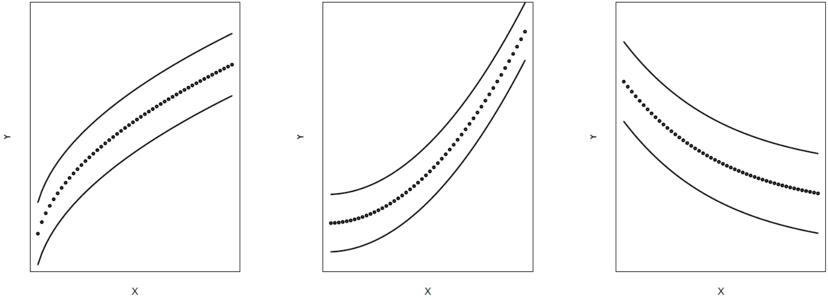
Prototypes for Transforming Y:



Try $\sqrt{Y}$, $\log_{10} Y$, or $1/Y$

Transformations - Prototypes for Transforming X

Prototypes for Transforming X:



Use $\sqrt{X}$ or $\log_{10} X$ (left); X^2 or $\exp(X)$ (middle); $1/X$ or $\exp(-X)$ (right).

Polynomial Regression Models

Recall that a **linear model** is defined as a model that is linear in the β's.

Consider the situation in which the analyst wants to create a model for the relationship between Y and X, but the relationship is polynomial in nature. For example

$$E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2.$$

This is an example of a **linear model**.

Another example of a linear model is one which contains **interaction** among a pair of predictor variables:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

In some applications there is a need to perform **transformations** on the predictor variables. For example,

$$E(Y) = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3.$$

Polynomial Regression Models - One Predictor, kth order

The kth order polynomial regression model with one predictor variable is defined by

$$E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots \beta_k x^k.$$

Generally, we start with a lower order model and try to add terms to possibly get a better model. But don't get carried away. Do not start with too high a power in a polynomial.

Polynomial Regression Models - Two Predictors

Consider a second-order polynomial regression model with two predictor variables

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2.$$

The coefficient β_{12} is often called the **interaction effect coefficient**.

The word **interaction** reflects the fact that a marginal change in Y when one predictor variable increases in value (while other **independent** variables are fixed) depends on the value of another **independent** variable.

The second-order no-interaction model with two predictor variables:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2.$$

Polynomial Regression Models - Interpreting the Response Surface

- When considered as a function of the regressors, the response surface is defined by the functional relationship

$$E(Y \mid X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) =$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k.$$

  If it is possible for the X_i to simultaneously take the value 0, then β_0 is the value of the response surface when all X_i equal 0. Otherwise, β_0 has no separate interpretation of its own.

- For $i = 1, \ldots, k$, β_i is interpreted as the change in the expected response per unit change in the regressor X_i, when all other regressors are held constant.

  Note that sometimes this interpretation is impossible. For example, if $X_1 = Z_1$ and $X_2 = Z_1^3$, we cannot change X_2 while holding X_1 constant.

- As a function of the **predictors** Z_i**s**, the response surface is defined by the functional relationship

$$E(Y \mid Z_1 = z_1, Z_2 = z_2, \ldots, Z_p = z_p) =$$

$$\beta_0 + \beta_1 X_1(z_1, z_2, \ldots, z_p) +$$

$$\beta_2 X_2(z_1, z_2, \ldots, z_p) +$$

$$\ldots + \beta_k X_k(z_1, z_2, \ldots, z_p)$$

Polynomial Regression Models - Interpreting the Response Surface

If the regressors are differentiable functions of the predictors, the instantaneous rate of change of the surface in the direction of predictor Z_i, at the point $z_1, z_2, \ldots, z_p$ is

$$\frac{\partial}{\partial z_i} E(Y \mid Z_1 = z_1, Z_2 = z_2, \ldots, Z_p = z_p).$$

(1) **Additive Model:** For the model

$$E(Y \mid Z_1 = z_1, Z_2 = z_2) = \beta_0 + \beta_1 z_1 + \beta_2 z_2,$$

the change in expected response per unit change in z_i is

$$\frac{\partial}{\partial z_i} E(Y \mid Z_1 = z_1, Z_2 = z_2) =$$

$$\frac{\partial}{\partial z_i}(\beta_0 + \beta_1 z_1 + \beta_2 z_2) = \beta_i,\ i = 1, 2.$$

(2) **Two predictor interaction Model:** For the two predictor interaction model

$$E(Y \mid Z_1 = z_1, Z_2 = z_2) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2,$$

the change in expected response per unit change in z_1 is

$$\frac{\partial}{\partial z_1} E(Y \mid Z_1 = z_1, Z_2 = z_2) = \beta_1 + \beta_3 z_2,$$

Polynomial Regression Models - Interpreting the Response Surface

(3) Full Quadratic Model: For the full quadratic model

$$E(Y \mid Z_1 = z_1, Z_2 = z_2) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1^2 + \beta_4 z_2^2 + \beta_5 z_1 z_2,$$

the change in expected response per unit change in z_1 is

$$\frac{\partial}{\partial z_1} E(Y \mid Z_1 = z_1, Z_2 = z_2) = \beta_1 + 2\beta_3 z_1 + \beta_5 z_2,$$

and the change in expected response per unit change in z_2 is

$$\frac{\partial}{\partial z_2} E(Y \mid Z_1 = z_1, Z_2 = z_2) = \beta_2 + 2\beta_4 z_2 + \beta_5 z_1.$$

Polynomial Regression Models - Hierarchical Models

Hierarchical Approach to Fitting: When using a polynomial regression model as an approximation to the true regression function, statisticians will often fit a second-order or third-order model and then explore whether a lower-order model is adequate.

- With the **hierarchical approach**, if a polynomial term of a given order is retained, then all related terms of lower order are also retained in the model. For example, all possible hierarchical models with two predictors are as the following

  The first-order model:
  $$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$
  The second-order no-interaction model:
  $$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$
  The model with first-order predictors and interaction:
  $$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$
  The complete second-order or full quadratic model
  $$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

- We'll discuss Polynomial Regression Models further later in the machine learning part.

Qualitative Predictors

Qualitative or categorical predictor variables can be used in regression models. Many predictor variables of interest in business, economics, and the social and biological sciences are categorical. Examples of categorical predictor variables are gender (male, female), purchase status (purchase, no purchase), and disability status (not disabled, partly disabled, fully disabled).

Example. Suppose we want to model Y (person's weight) as a function of X_1 (person's height) and X_2 (Gender), where

$$X_2 = \begin{cases} 1 & \text{Male} \\ 0 & \text{Female} \end{cases}$$

Consider the model

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

For males, it becomes

$$E(Y) = (\beta_0 + \beta_2) + \beta_1 x_1.$$

For females, it becomes

$$E(Y) = \beta_0 + \beta_1 x_1.$$

Qualitative Predictors

We may not believe that these lines are parallel. If so, you should introduce an interaction term into the model:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

For males, it becomes

$$E(Y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1.$$

For females, it becomes

$$E(Y) = \beta_0 + \beta_1 x_1.$$

Why do we combine the data among men and women? Why do not we just model them separately?

Suppose we observe n_1 males and n_2 females.

$$\begin{aligned} \text{males SSE has df } &= n_1 - 2 \\ \text{females SSE has df } &= n_2 - 2 \end{aligned}$$

But when we combine men and women using the model with interaction effect,

$$\text{SSE has df } = n_1 + n_2 - 3.$$

The larger df is an advantage as long as $\sigma_M^2 = \sigma_F^2$

Categorical in Regression - Several Categories

If there are c categories, we need $c-1$ categorical variables.

$$Z_1 = \begin{cases} 1 & \text{category level 1} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_2 = \begin{cases} 1 & \text{category level 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$Z_{c-1} = \begin{cases} 1 & \text{category level } c-1 \\ 0 & \text{otherwise} \end{cases}$$

Example

Example. Table 8.2 An economist wished to relate the speed with which a particular insurance innovation is adopted (Y) to the size of the insurance firm (X_1) and the type of firm (X_2). The dependent variable is measured by the number of months elapsed. between the time the first firm adopted the innovation and the time the given firm adopted the innovation. The first independent variable, size of firm, is quantitative, and is measured by the amount of total assets of the firm. The second independent variable, type of firm, is qualitative and is composed of two classes-stock companies and mutual companies.

Next slide shows the complete data.

Example

```
data ch8tab02;
input y x1 x2@@;
label y = 'Months'
x1 = 'Size'
x2 = 'Firm Indicator';
cards;
17 151 0 26 92 0
21 175 0 30 31 0
22 104 0 0 277 0
12 210 0 19 120 0
4 290 0 16 238 0
28 164 1 15 272 1
11 295 1 38 68 1
31 85 1 21 224 1
20 166 1 13 305 1
30 124 1 14 246 1
;
run;
```

Example

```
proc reg data = ch8tab02;
model y = x1 x2/ clb;
run;
```

Example

```
data ch8tab02;
set ch8tab02;
if x2 = 0 then do;
z1 = x1;
y1 = y;
end;
if x2= 1 then do;
z2 = x1;
y2 = y;
end;
run;
```

```
proc reg data = ch8tab02;
model y = z1;
run;
```

```
proc reg data = ch8tab02;
model y = z2;
run;
```

License