Remedial Measures Wrap-Up
and Transformations – Box Cox

Dr. Frank Wood
Last Class

• Graphical procedures for determining appropriateness of regression fit
  – Normal probability plot
• Tests to determine
  – Constancy of error variance
  – Appropriateness of linear fit
• What do we if we determine (through testing or otherwise) that the linear regression fit is not good?
Overview of Remedial Measures

• If simple regression model is not appropriate there are two choices
  1. Abandon simple regression model and develop and use a more appropriate model
  2. Employ some transformation of the data so that the simple regression model is appropriate for the transformed data.
Fixes For…

- Nonlinearity of regression function
  - Transformation(s) (today)
- Nonconstancy of error variance
  - Weighted least squares (nice project idea, coming later in class) and transformations
- Nonindependence of error terms
  - Directly model correlation or use first differences (may skip)
- Nonnormality of error terms
  - Transformation(s) (today)
- Omission of important predictor variables
  - Multiple regression – coming soon
- Outlying observations
  - Robust regression (another nice project idea)
Nonlinearity of regression function

• Direct approach

  • Modify regression model by altering the nature of the regression function. For instance a quadratic regression function might be used

\[ E\{Y\} = \beta_0 + \beta_1 X + \beta_2 X^2 \]

  • or an exponential function

\[ E\{Y\} = \beta_0 \beta_1^X \]

  • Such approaches employ a transformation to (approximately) linearize a regression function
Quick Questions

- How would you fit such models?
- How does the exponential regression function relate to regular linear regression?
- Where did the error terms go?
Transformations

• Transformations for Nonlinear Relation Only
  – Appropriate when the distribution of the error terms is reasonably close to a normal distribution
  – In this situation
    • transformation of X should be attempted
    • transformation of Y should not be attempted because it will materially effect the distribution of the error terms
Prototype Regression Patterns

<table>
<thead>
<tr>
<th>Prototype Regression Pattern</th>
<th>Transformations of $X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$X' = \log_{10} X$    $X' = \sqrt{X}$</td>
</tr>
<tr>
<td>(b)</td>
<td>$X' = X^2$            $X' = \exp(X)$</td>
</tr>
<tr>
<td>(c)</td>
<td>$X' = 1/X$            $X' = \exp(-X)$</td>
</tr>
</tbody>
</table>

Note constancy error terms
Example

- Experiment
  - $X$: days of training received
  - $Y$: sales performance (score)
$X' = \sqrt{X}$
### Example Data Transformation

<table>
<thead>
<tr>
<th>Sales Trainee</th>
<th>(1) Days of Training</th>
<th>(2) Performance Score</th>
<th>(3) $X'_i = \sqrt{X_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>$X_i$</td>
<td>$Y_i$</td>
<td>$X'_i$</td>
</tr>
<tr>
<td>1</td>
<td>.5</td>
<td>42.5</td>
<td>.70711</td>
</tr>
<tr>
<td>2</td>
<td>.5</td>
<td>50.6</td>
<td>.70711</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>68.5</td>
<td>1.00000</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>80.7</td>
<td>1.00000</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>89.0</td>
<td>1.22474</td>
</tr>
<tr>
<td>6</td>
<td>1.5</td>
<td>99.6</td>
<td>1.22474</td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>105.3</td>
<td>1.41421</td>
</tr>
<tr>
<td>8</td>
<td>2.0</td>
<td>111.8</td>
<td>1.41421</td>
</tr>
<tr>
<td>9</td>
<td>2.5</td>
<td>112.3</td>
<td>1.58114</td>
</tr>
<tr>
<td>10</td>
<td>2.5</td>
<td>125.7</td>
<td>1.58114</td>
</tr>
</tbody>
</table>

\[
\hat{Y} = -10.33 + 83.45X'
\]
Graphical Residual Analysis

(a) Scatter Plot

(b) Scatter Plot against $\sqrt{X}$

(c) Residual Plot against $\sqrt{X}$

(d) Normal Probability Plot
Matlab

• Run matlab_demos\transform_X.m
Transformations on $Y$

- Nonnormality and unequal variances of error terms frequently appear together
- To remedy these in the normal regression model we need a transformation on $Y$
- This is because
  - Shapes and spreads of distributions of $Y$ need to be changed
  - May help linearize a curvilinear regression relation
- Can be combined with transformation on $X$
Prototype Regression Patterns and Y Transformations

Prototype Regression Pattern

Transformations on $Y$

$Y' = \sqrt{Y}$

$Y' = \log_{10} Y$

$Y' = \frac{1}{Y}$

Note change in error distribution as function of input
Example

• Use of logarithmic transformation of $Y$ to linearize regression relations and stabilize error variance

• Data on age ($X$) and plasma level of a polyamine ($Y$) for a portion of the 25 healthy children in a study. Younger children exhibit greater variability than older children.
Plasma level versus age

(a) Scatter Plot

(b) Scatter Plot with $Y' = \log_{10} Y$

(c) Residual Plot against $X$

(d) Normal Probability Plot
### Associated Data

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
<th>Plasma Level</th>
<th>Y_i' = \log_{10} Y_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (newborn)</td>
<td>13.44</td>
<td>1.1284</td>
</tr>
<tr>
<td>2</td>
<td>0 (newborn)</td>
<td>12.84</td>
<td>1.1086</td>
</tr>
<tr>
<td>3</td>
<td>0 (newborn)</td>
<td>11.91</td>
<td>1.0759</td>
</tr>
<tr>
<td>4</td>
<td>0 (newborn)</td>
<td>20.09</td>
<td>1.3030</td>
</tr>
<tr>
<td>5</td>
<td>0 (newborn)</td>
<td>15.60</td>
<td>1.1931</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>10.11</td>
<td>1.0048</td>
</tr>
<tr>
<td>7</td>
<td>1.0</td>
<td>11.38</td>
<td>1.0561</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>19</td>
<td>3.0</td>
<td>6.90</td>
<td>0.8388</td>
</tr>
<tr>
<td>20</td>
<td>3.0</td>
<td>6.77</td>
<td>0.8306</td>
</tr>
<tr>
<td>21</td>
<td>4.0</td>
<td>4.86</td>
<td>0.6866</td>
</tr>
<tr>
<td>22</td>
<td>4.0</td>
<td>5.10</td>
<td>0.7076</td>
</tr>
<tr>
<td>23</td>
<td>4.0</td>
<td>5.67</td>
<td>0.7536</td>
</tr>
<tr>
<td>24</td>
<td>4.0</td>
<td>5.75</td>
<td>0.7597</td>
</tr>
<tr>
<td>25</td>
<td>4.0</td>
<td>6.23</td>
<td>0.7945</td>
</tr>
</tbody>
</table>
• If we fit a simple linear regression line to the log transformed Y data we obtain

\[ \hat{Y}' = 1.135 - .1023X \]

• And the coefficient of correlation between the ordered residuals and their expected values under normality is .981 (for \( \alpha = .05 \) B.6 in the book shows a critical value of .959)

• Normality of error terms supported, regression model for transformed Y data appropriate
Box Cox Transforms

• It can be difficult to graphically determine which transformation of Y is most appropriate for correcting
  – skewness of the distributions of error terms
  – unequal variances
  – nonlinearity of the regression function

• The Box-Cox procedure automatically identifies a transformation from the family of power transformations on Y
Box Cox Transforms

- This family is of the form

\[ Y' = Y^\lambda \]

- Examples include

\[
\begin{align*}
\lambda = 2 & \quad Y' = Y^2 \\
\lambda = .5 & \quad Y' = \sqrt{Y} \\
\lambda = 0 & \quad Y' = \log_e Y \quad \text{(by definition)} \\
\lambda = -.5 & \quad Y' = \frac{1}{\sqrt{Y}} \\
\lambda = -1.0 & \quad Y' = \frac{1}{Y}
\end{align*}
\]
Box Cox Cont.

- The normal error regression model with the response variable a member of the family of power transformations becomes

\[ Y_i^\lambda = \beta_0 + \beta_1 X_i + \varepsilon_i \]

- This model has an additional parameter that needs to be estimated
- Maximum likelihood is a way to estimate this parameter
Box Cox Maximum Likelihood Estimation

- Before setting up maximum likelihood estimation, the observations are further standardized so that the magnitude of the error sum of squares does not depend on the value of $\lambda$

- The transformation is given by

$$W_i = \begin{cases} 
K_1(Y_i^\lambda - 1) & \lambda \neq 0 \\
K_2(\log_e Y_i) & \lambda = 0
\end{cases}$$

where

$$K_2 = \left(\prod_{i=1}^{n} Y_i\right)^{1/n} \quad \text{geometric mean}$$

$$K_1 = \frac{1}{\lambda K_2^{\lambda - 1}}$$
Box Cox Maximum Likelihood Estimation

- Maximize

\[ \log(L(X, Y, \sigma, \lambda, b_1, b_0)) = -\sum_i \frac{(W_i - (b_1 X_i + b_0))^2}{2\sigma^2} - n\log(\sigma) \]

w.r.t. \( \lambda, \sigma, b_1, \) and \( b_0 \)

- How?
  - Take partial derivatives
  - Solve
  - or... gradient ascent methods

Show box_cox_demo.m
Comments on Box Cox

- The Box-Cox procedure is ordinarily used only to provide a guide for selecting a transformation.
- At times, theoretical or other a priori considerations can be utilized to help in choosing an appropriate transformation.
- It is important to perform residual analysis after the transformation to ensure that the transformation is appropriate.
- When transformed models are employed, $b_0$ and $b_1$ obtained via least squares have the least squares property w.r.t. the transformed observations not the original ones.