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

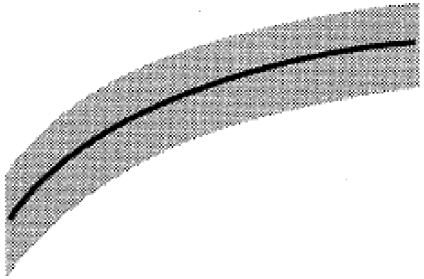
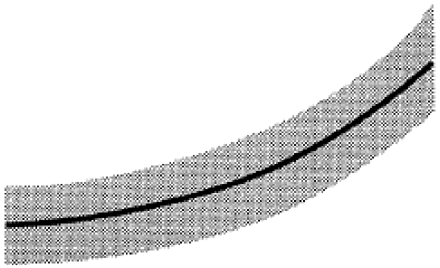
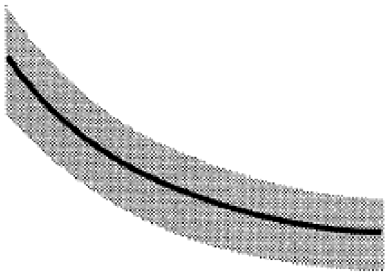
Quick Questions

- 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

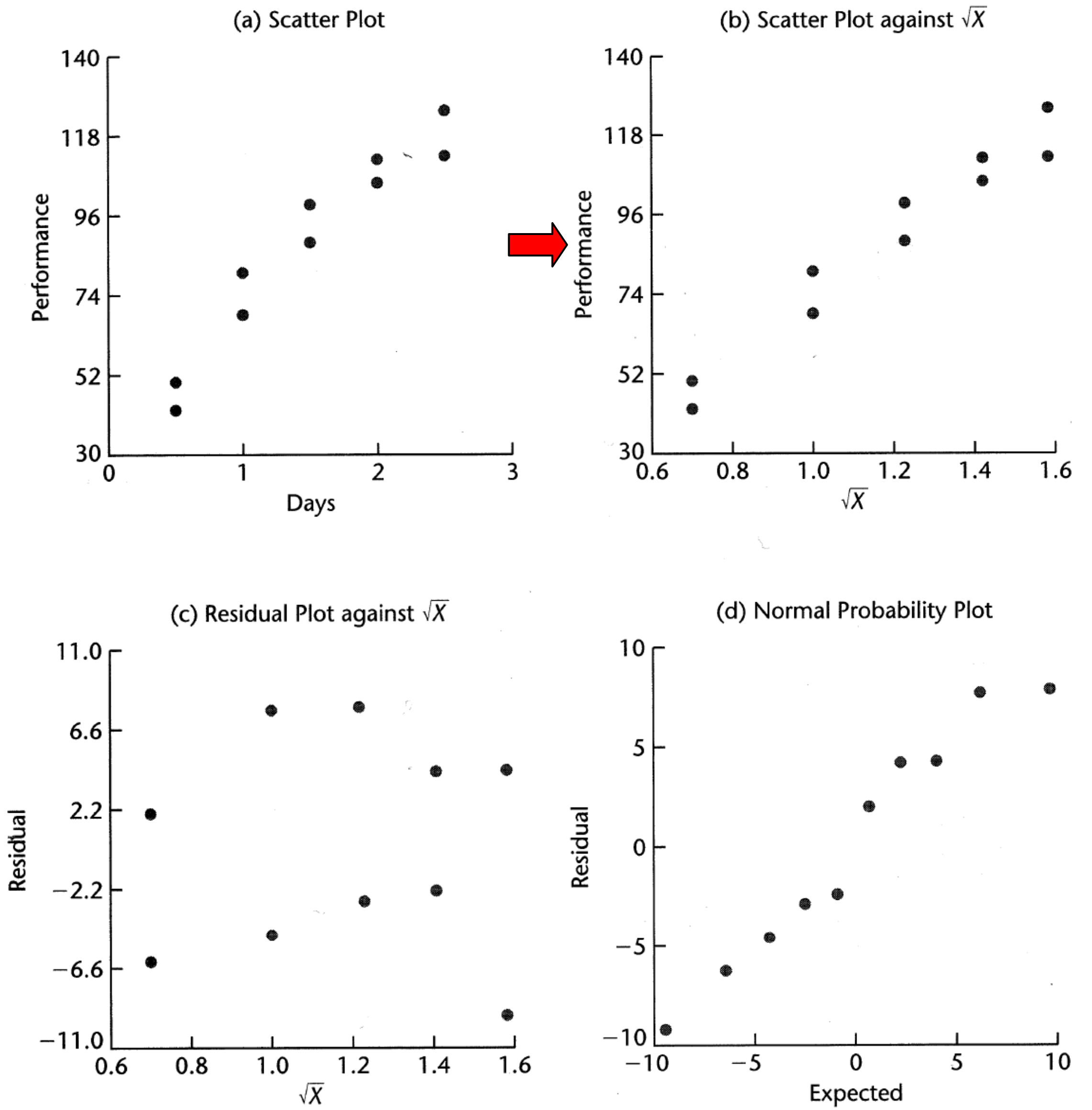
	Prototype Regression Pattern	Transformations of X
(a)		$X' = \log_{10} X$ $X' = \sqrt{X}$
(b)		$X' = X^2$ $X' = \exp(X)$
(c)		$X' = 1/X$ $X' = \exp(-X)$

Note constancy error terms

Example

- Experiment
  - X : days of training received
  - Y : sales performance (score)

X’ = sqrt(X)



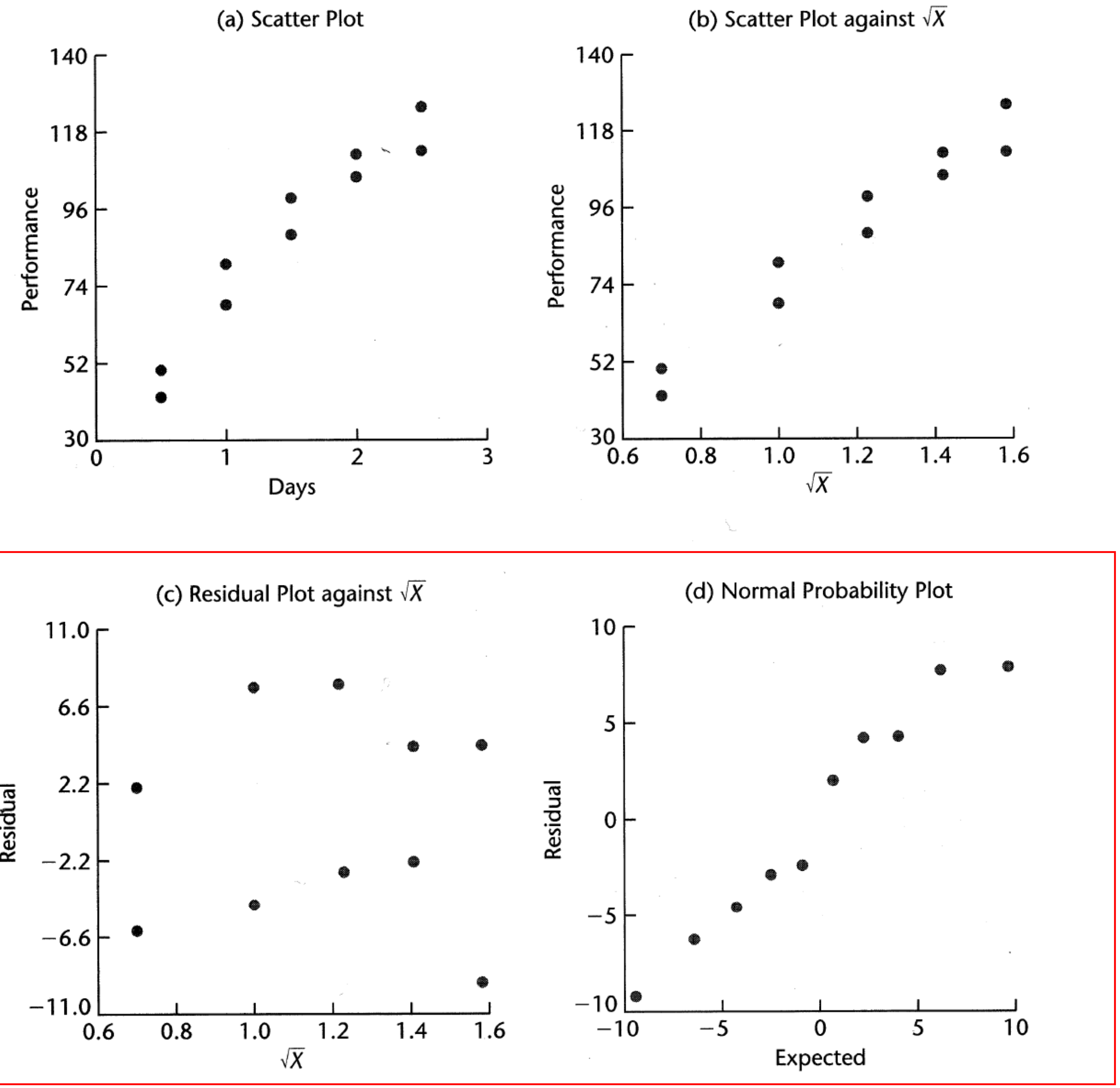
(a) Scatter Plot
Performance
140
118
96
74
52
30
0
1
2
3
Days
(b) Scatter Plot against √X
Performance
140
118
96
74
52
30
0.6
0.8
1.0
1.2
1.4
1.6
√X
(c) Residual Plot against √X
Residual
11.0
6.6
2.2
−2.2
−6.6
−11.0
0.6
0.8
1.0
1.2
1.4
1.6
√X
(d) Normal Probability Plot
Residual
10
5
0
−5
−10
−10
−5
0
5
10
Expected

Example Data Transformation

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

$$\hat{Y} = -10.33 + 83.45X'$$

Graphical Residual Analysis



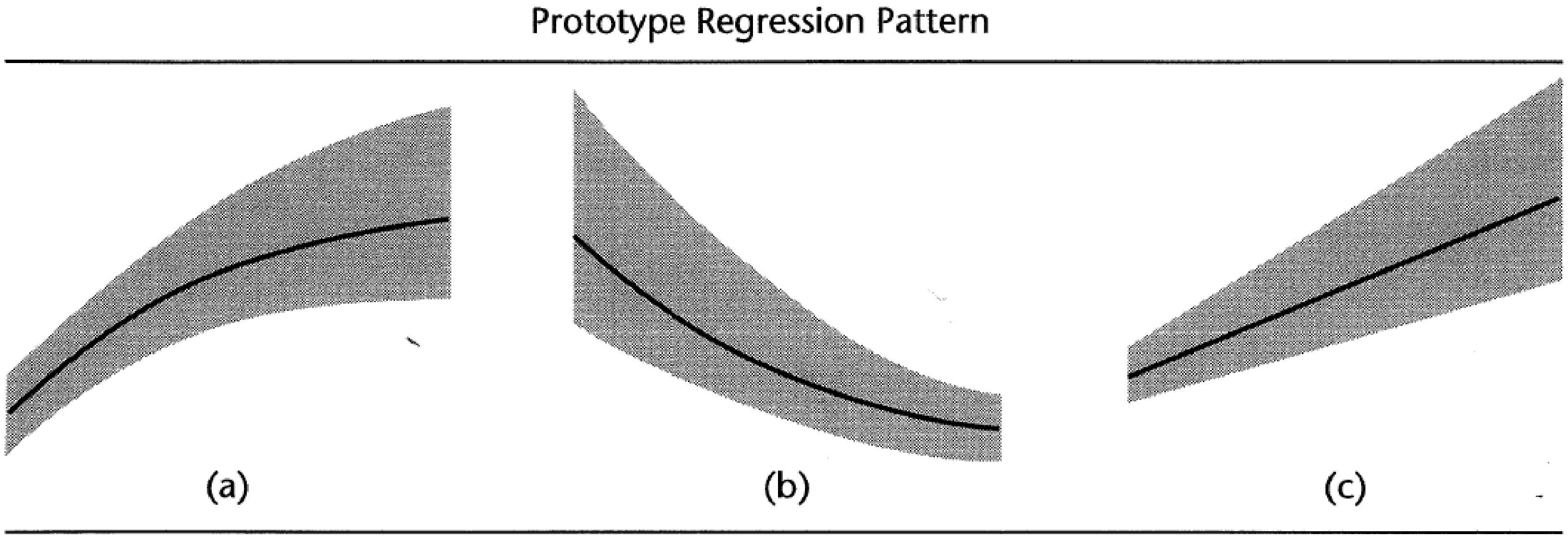
Matlab

- 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



Note change in
error distribution
as function of input

Transformations on Y

$$Y' = \sqrt{Y}$$

$$Y' = \log_{10} Y$$

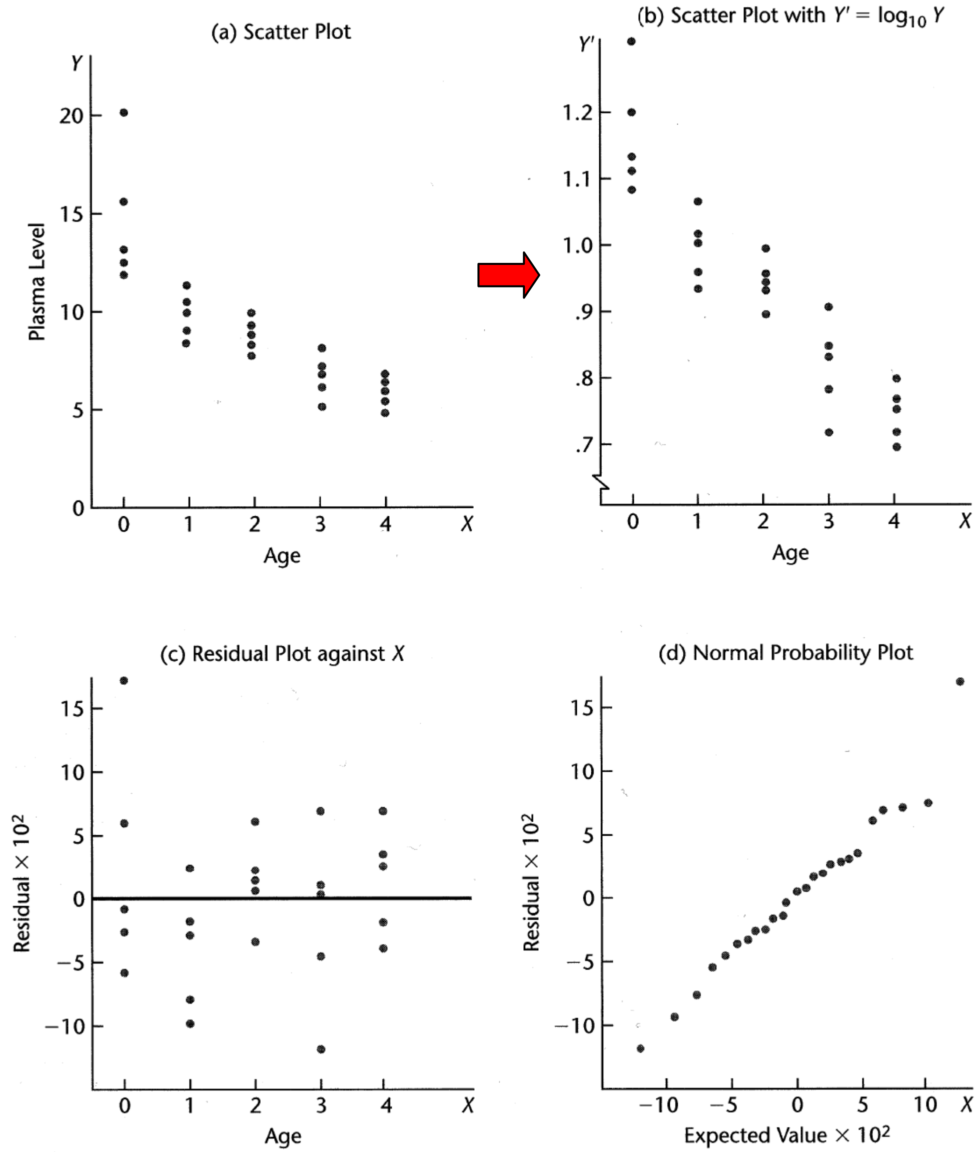
$$Y' = 1/Y$$

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



Associated Data

Child i	(1) Age X_i	(2) Plasma Level Y_i	(3) $Y_i' = \log_{10} Y_i$
1	0 (newborn)	13.44	1.1284
2	0 (newborn)	12.84	1.1086
3	0 (newborn)	11.91	1.0759
4	0 (newborn)	20.09	1.3030
5	0 (newborn)	15.60	1.1931
6	1.0	10.11	1.0048
7	1.0	11.38	1.0561
...	...	...	...
19	3.0	6.90	.8388
20	3.0	6.77	.8306
21	4.0	4.86	.6866
22	4.0	5.10	.7076
23	4.0	5.67	.7536
24	4.0	5.75	.7597
25	4.0	6.23	.7945

- 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 α = .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

$$\begin{array}{lll}
\lambda = 2 & Y' = Y^2 & \\
\lambda = .5 & Y' = \sqrt{Y} & \\
\lambda = 0 & Y' = \log_e Y & \text{(by definition)} \\
\lambda = -.5 & Y' = \dfrac{1}{\sqrt{Y}} & \\
\lambda = -1.0 & Y' = \dfrac{1}{Y} &
\end{array}$$

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 λ
- The transformation is given by

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

where

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

$$K_1 = \frac{1}{\lambda K_2^{\lambda-1}}$$

Box Cox Maximum Likelihood Estimation

- Maximize

$$log(L(X,Y,\sigma,\lambda,b_1,b_0)) = -\sum_i \frac{(W_i-(b_1X_i+b_0))^2}{2\sigma^2} - nlog(\sigma)$$

  w.r.t. λ, σ, 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.