Applied Regression Modeling: A Business Approach
Chapter 4: Regression Model Building I
Section 4.1

by Iain Pardoe
A *transformation* is a mathematical function applied to a variable in our dataset.

Example: it is possible that there is a stronger relationship between $Y$ and $\log_e(X)$ than between $Y$ and $X$.

To find out, fit two models:

- Model 1: $E(Y) = b_0 + b_1 X$;
- Model 2: $E(Y) = b_0 + b_1 \log_e(X)$.

Which model fits better?

- Smaller regression standard error, $s$;
- Larger coefficient of determination, $R^2$;
- Larger magnitude individual t-statistic.
TV commercial data: $X =$ spending in $m$, $Y =$ millions of retained impressions.

<table>
<thead>
<tr>
<th>Firm</th>
<th>$X$ (spending)</th>
<th>$Y$ (retained impressions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miller Lite</td>
<td>50.1</td>
<td>32.1</td>
</tr>
<tr>
<td>Pepsi</td>
<td>74.1</td>
<td>99.6</td>
</tr>
<tr>
<td>Stroh’s</td>
<td>19.3</td>
<td>11.7</td>
</tr>
<tr>
<td>Federal Express</td>
<td>22.9</td>
<td>21.9</td>
</tr>
<tr>
<td>Burger King</td>
<td>82.4</td>
<td>60.8</td>
</tr>
<tr>
<td>Coca-Cola</td>
<td>40.1</td>
<td>78.6</td>
</tr>
<tr>
<td>McDonald’s</td>
<td>185.9</td>
<td>92.4</td>
</tr>
<tr>
<td>MCI</td>
<td>26.9</td>
<td>50.7</td>
</tr>
<tr>
<td>Diet-Cola</td>
<td>20.4</td>
<td>21.4</td>
</tr>
<tr>
<td>Ford</td>
<td>166.2</td>
<td>40.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Kibbles ’n Bits</td>
<td>6.1</td>
<td>4.4</td>
</tr>
</tbody>
</table>
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

TVADS example

Scatterplots for models 1 and 2

Regression results for models 1 and 2

Interpretation

Why it works

Selecting transformations

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

4.1.5 Transformations for the response and predictors

Model 1 on the left: \( E(Y) = b_0 + b_1 X \).

Model 2 on the right: \( E(Y) = b_0 + b_1 \log_e(X) \).

Plots include fitted regression (least squares) lines.

Which model fits the data better?
### Regression results for models 1 and 2

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.651&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.424</td>
<td>0.394</td>
<td>23.5015</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Intercept), X.

#### Parameters<sup>a</sup>

| Model | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------|----------|------------|--------|---------|
| 1 (Intercept) | 22.163 | 7.089 | 3.126 | 0.006 |
| X       | 0.363    | 0.097 | 3.739 | 0.001 |

<sup>a</sup> Response variable: Y.

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.730&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.532</td>
<td>0.508</td>
<td>21.1757</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Intercept), logX.

#### Parameters<sup>a</sup>

| Model | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------|----------|------------|--------|---------|
| 2 (Intercept) | -28.050 | 15.441 | -1.817 | 0.085 |
| logX   | 20.180   | 4.339     | 4.650  | 0.000  |

<sup>a</sup> Response variable: Y.
4.1 Transformations
4.1.1 Natural logarithm transformation for predictors
TVADS example
Scatterplots for models 1 and 2
Regression results for models 1 and 2

Interpretation
Why it works
Selecting transformations
4.1.2 Polynomial transformation for predictors
4.1.3 Reciprocal transformation for predictors
4.1.4 Natural logarithm transformation for the response
4.1.5 Transformations for the response and predictors

- Model 2 with $\log_e(X)$ more usefully describes relationship between TV commercial success and cost than model 1 with untransformed $X$:
  - smaller $s$, larger $R^2$, larger magnitude individual t-statistic.
- However, model interpretation less straightforward.
- For example, as $X$ increases from 10 to 20, expect $Y$ to increase from $-28.1 + 20.2 \log_e(10) = 18.4$ to $-28.1 + 20.2 \log_e(20) = 32.4$.
- But as $X$ increases from 150 to 160, expect $Y$ to increase from $-28.1 + 20.2 \log_e(150) = 73.1$ to $-28.1 + 20.2 \log_e(160) = 74.4$.
- Predictor effect plot (Section 5.4) displays this graphically.
Multiple linear regression models are often more effective when predictors have reasonably symmetric distributions and are not too highly skewed.

Natural log. transformation works well for positively skewed variables with a few values much higher than the majority, since it tends to spread out lower values and pull in the higher values (i.e. makes distribution more symmetric).
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

TVADS example
Scatterplots for models 1 and 2
Regression results for models 1 and 2
Interpretation
Why it works

Selecting transformations

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

4.1.5 Transformations for the response and predictors

Transformations often suggested by theories about economics, consumer psychology, worker behavior, business decision-making, and so on.

At other times, we might observe particular empirical relationships in sample datasets, and we can try out various variable transformations to see how to best model the data.

Common transformations in business:

- natural logarithm, e.g., $\log_e(X)$;
- polynomial, e.g., $X$, $X^2$, $X^3$, ...;
- reciprocal, e.g., $1/X$;
- square root, e.g., $\sqrt{X}$. 

© Iain Pardoe, 2006
• Investigate whether the age of a home factors into its sale price in a particular housing market.
• For the sake of illustration, ignore other predictors (floor size, lot size, etc.) and focus solely on $X = \text{age}$, defined as 2005 minus year built.
• Realtor experience suggests both older and newer homes command a price premium relative to “middle-aged” homes in this market.
• Which of these models can capture such a trend?
  - Model 1: $\mathbb{E}(Y) = b_0 + b_1 X$
  - Model 2: $\mathbb{E}(Y) = b_0 + b_1 X + b_2 X^2$. 
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors HOMES4 example

Scatterplot with model 1 and 2 fitted lines

Regression results for models 1 and 2 Interpretation Polynomial transformations in practice

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

4.1.5 Transformations for the response and predictors

Model 1: $E(Y) = b_0 + b_1X$ (straight)
Model 2: $E(Y) = b_0 + b_1X + b_2X^2$ (curved).

Which model fits the data better?
Regression results for models 1 and 2

### Parameters \(^a\)

| Model | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------|----------|------------|--------|----------|
| 1     |          |            |        |          |
| (Intercept) | 299.883  | 12.554     | 23.887 | 0.000    |
| X     | −0.396   | 0.295      | −1.342 | 0.184    |

\(^a\) Response variable: Y.

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.383 (^a)</td>
<td>0.147</td>
<td>0.123</td>
<td>56.486</td>
</tr>
</tbody>
</table>

\(^a\) Predictors: (Intercept), X, Xsq.

### Parameters \(^a\)

| Model | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------|----------|------------|--------|----------|
| 2     |          |            |        |          |
| (Intercept) | 330.407  | 15.103     | 21.876 | 0.000    |
| X     | −2.652   | 0.749      | −3.542 | 0.001    |
| Xsq   | 0.027    | 0.008      | 3.245  | 0.002    |

\(^a\) Response variable: Y.
Model 2 with $X$ and $X^2$ more usefully describes relationship between sale price and age than model 1 with untransformed $X$:

- two tail p-value for testing $b_2$ in model 2 is 0.002.

However, model interpretation less straightforward.

For example, $Y$ decreases quite steeply as $X$ increases between 0 and 20, levels off for $X$ between 20 and 70, and then increases more steeply as $X$ increases between 70 and 100.

Quick calculations can quantify these changes more precisely.

Predictor effect plot (Section 5.4) displays this graphically.
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

HOMES4 example
Scatterplot with model 1 and 2 fitted lines
Regression results for models 1 and 2
Interpretation

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

4.1.5 Transformations for the response and predictors

© Iain Pardoe, 2006

- General polynomial model:
  \[ E(Y) = b_0 + b_1 X + b_2 X^2 + b_3 X^3 + \cdots \]
- Rare to see powers higher than two (quadratic) or three (cubic) in linear regression models unless theoretical reasons for including higher powers.
Polynomial transformations in practice

4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

HOMES4 example
Scatterplot with model 1 and 2 fitted lines
Regression results for models 1 and 2
Interpretation

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

4.1.5 Transformations for the response and predictors

© Iain Pardoe, 2006

- General polynomial model:
  \[ E(Y) = b_0 + b_1 X + b_2 X^2 + b_3 X^3 + \cdots. \]

- Rare to see powers higher than two (quadratic) or three (cubic) in linear regression models unless theoretical reasons for including higher powers.

- When using \( X^2 \), \( X^3 \), etc., in models, lower powers are often included, regardless of their significance—this is called preserving hierarchy.
  - e.g., keep \( X \) in the model if \( X^2 \) is significant (has a small p-value);
  - e.g., keep \( X \) and \( X^2 \) in the model if \( X^3 \) is significant (has a small p-value).
Polynomial transformations in practice

- General polynomial model:
  \[ E(Y) = b_0 + b_1 X + b_2 X^2 + b_3 X^3 + \cdots. \]
- Rare to see powers higher than two (quadratic) or three (cubic) in linear regression models unless theoretical reasons for including higher powers.
- When using \( X^2, X^3, \) etc., in models, lower powers are often included, *regardless* of their significance—this is called preserving *hierarchy*.
  - e.g., keep \( X \) in the model if \( X^2 \) is significant (has a small p-value);
  - e.g., keep \( X \) and \( X^2 \) in the model if \( X^3 \) is significant (has a small p-value).
- When using \( X^2, X^3, \) etc., in models, common to first *rescale* values of \( X \) to have mean \( \approx 0 \) and std. dev. \( \approx 1 \) (example in Section 6.1).
• Investigate any relationship between a car’s city miles per gallon \((Y)\) and its weight \((X)\).
• For the sake of illustration, ignore other predictors (horsepower, engine size, etc.).
• Engineering principals suggest that there is an “inverse” relationship between weight and fuel efficiency.
• Which of these models can capture such a trend?
  ○ Model 1 : \(E(Y) = b_0 + b_1 X;\)
  ○ Model 2 : \(E(Y) = b_0 + b_1 (1/X).\)
Model 1 on the left: $E(Y) = b_0 + b_1 X$.
Model 2 on the right: $E(Y) = b_0 + b_1 (1/X)$.

Plots include fitted regression (least squares) lines. Which model fits the data better?
Regression results for models 1 and 2

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.837&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.700</td>
<td>0.700</td>
<td>2.291</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Intercept), X.

### Parameters<sup>a</sup>

| Model   | Estimate | Std. Error | t-stat | Pr(>|t|) |
|---------|----------|------------|--------|---------|
| 1 (Intercept) | 37.020   | 0.572      | 64.760 | 0.000   |
| X       | -4.809   | 0.157      | -30.723| 0.000   |

<sup>a</sup> Response variable: Y.

---

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.895&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.800</td>
<td>0.800</td>
<td>1.869</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Intercept), recipX.

### Parameters<sup>a</sup>

| Model   | Estimate | Std. Error | t-stat | Pr(>|t|) |
|---------|----------|------------|--------|---------|
| 2 (Intercept) | 1.195    | 0.472      | 2.534  | 0.012   |
| recipX  | 64.019   | 1.590      | 40.251 | 0.000   |

<sup>a</sup> Response variable: Y.
• Model 2 with $1/X$ more usefully describes relationship between city miles per gallon and weight than model 1 with untransformed $X$:
  - smaller $s$, larger $R^2$, larger magnitude individual $t$-statistic.

• However, model interpretation less straightforward.
• For example, as $X$ increases from 2 to 3, expect $Y$ to decrease from $1.2 + 64.0(1/2) = 33.2$ to $1.2 + 64.0(1/3) = 22.5$.
• But as $X$ increases from 5 to 6, expect $Y$ to decrease from $1.2 + 64.0(1/5) = 14.0$ to $1.2 + 64.0(1/6) = 11.9$.
• Predictor effect plot (Section 5.4) displays this graphically.
Investigate any relationship between a worker’s salary ($Y$) and experience ($X$).

For the sake of illustration, ignore other predictors (job title, education, etc.).

Annual salary increases are usually proportional (% increases) rather than additive ($\$ increases). This suggests that there should be some kind of “multiplicative” relationship between additional years of experience and salary differences.

Which of these models can capture such a trend?

- Model 1: $E(Y) = b_0 + b_1 X$
- Model 2: $E(\log_e(Y)) = b_0 + b_1 X$. 
Multiplicative models

4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

WORKEXP example

Multiplicative models

Scatterplots for models 1 and 2
Regression results for models 1 and 2
Interpretation

4.1.5 Transformations for the response and predictors

• Consider Model 1: \( E(Y) = b_0 + b_1 X \).
• What happens to \( Y \) when \( X \) increases one unit?
Consider Model 1: \( \text{E}(Y) = b_0 + b_1 X \).

What happens to \( Y \) when \( X \) increases one unit?

- \( Y_{\text{after}} - Y_{\text{before}} = \left[ b_0 + b_1 (X + 1) \right] - \left[ b_0 + b_1 X \right] = b_1 \).
• Consider Model 1: $E(Y) = b_0 + b_1 X$.
• What happens to $Y$ when $X$ increases one unit?
  \[ Y_{\text{after}} - Y_{\text{before}} = [b_0 + b_1(X+1)] - [b_0 + b_1 X] = b_1. \]
• Consider Model 2: $E(\log_e(Y)) = b_0 + b_1 X$.
• What happens to $Y$ when $X$ increases one unit?
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

WORKEXP example

Multiplicative models

Scatterplots for models 1 and 2
Regression results for models 1 and 2
Interpretation

4.1.5 Transformations for the response and predictors

- Consider Model 1: \( E(Y) = b_0 + b_1 X \).
- What happens to \( Y \) when \( X \) increases one unit?
  - \( Y_{\text{after}} - Y_{\text{before}} = \left[ b_0 + b_1(X + 1) \right] - \left[ b_0 + b_1X \right] = b_1 \).
- Consider Model 2: \( E(\log_e(Y)) = b_0 + b_1 X \).
- What happens to \( Y \) when \( X \) increases one unit?
  - \( Y_{\text{after}} - Y_{\text{before}} = \exp(b_0 + b_1(X + 1)) - \exp(b_0 + b_1X) \\
    = \left[ \exp(b_1) - 1 \right] \left[ \exp(b_0 + b_1X) \right]. \)
• Consider Model 1: \( E(Y) = b_0 + b_1 X \).

• What happens to \( Y \) when \( X \) increases one unit?
  ◦ \( Y_{\text{after}} - Y_{\text{before}} = [b_0 + b_1 (X + 1)] - [b_0 + b_1 X] = b_1 \).

• Consider Model 2: \( E(\log_e(Y)) = b_0 + b_1 X \).

• What happens to \( Y \) when \( X \) increases one unit?
  ◦ \( Y_{\text{after}} - Y_{\text{before}} \)
    \[= \exp(b_0 + b_1 (X + 1)) - \exp(b_0 + b_1 X)\]
    \[= [\exp(b_1) - 1][\exp(b_0 + b_1 X)].\]
  ◦ \( (Y_{\text{after}} - Y_{\text{before}})/Y_{\text{before}} = \exp(b_1) - 1.\)
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

WORKEXP example

- Consider Model 1: \( \mathbb{E}(Y) = b_0 + b_1 X \).
- What happens to \( Y \) when \( X \) increases one unit?
  \( Y_{\text{after}} - Y_{\text{before}} = [b_0 + b_1(X+1)] - [b_0 + b_1X] = b_1. \)
- Consider Model 2: \( \mathbb{E}(\log_e(Y)) = b_0 + b_1 X \).
- What happens to \( Y \) when \( X \) increases one unit?
  \[ Y_{\text{after}} - Y_{\text{before}} = \exp(b_0 + b_1(X+1)) - \exp(b_0 + b_1X) = [\exp(b_1) - 1][\exp(b_0 + b_1X)]. \]
  \[ (Y_{\text{after}} - Y_{\text{before}})/Y_{\text{before}} = \exp(b_1) - 1. \]
- In other words, in model 2, \( \exp(b_1) - 1 \) represents the proportional increase in salary (\( Y \)) when experience (\( X \)) increases by one year.
  - Example: if \( \exp(b_1) - 1 = 0.06 = 6\% \), then we expect annual salary increases to be 6\%, from $50,000 to $53,000, say ((53-50)/50 = 0.06).
Model 1 on the left: \( E(Y) = b_0 + b_1X \).
Model 2 on the right: \( E(\log_e(Y)) = b_0 + b_1X \).

Plots include fitted regression (least squares) lines. Which model fits the data better?
### Regression results for models 1 and 2

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.930 (^a)</td>
<td>0.865</td>
<td>0.862</td>
<td>8.357</td>
</tr>
</tbody>
</table>

\(^a\) Predictors: (Intercept), X.

#### Parameters \(^a\)

| Model | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------|----------|------------|--------|----------|
| 1     | (Intercept) | 10.323     | 2.706  | 3.815    | 0.000    |
|       | X         | 3.094      | 0.177  | 17.515   | 0.000    |

\(^a\) Response variable: Y.

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Regression Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.955 (^a)</td>
<td>0.912</td>
<td>0.910</td>
<td>0.125</td>
</tr>
</tbody>
</table>

\(^a\) Predictors: (Intercept), X.

#### Parameters \(^a\)

| Model | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------|----------|------------|--------|----------|
| 2     | (Intercept) | 3.074      | 0.040  | 76.089   | 0.000    |
|       | X         | 0.059      | 0.003  | 22.302   | 0.000    |

\(^a\) Response variable: logY.
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

WORKEXP example

Multiplicative models

Scatterplots for models 1 and 2

Regression results for models 1 and 2

4.1.5 Transformations for the response and predictors

• Model 2 with $\log_e(Y)$ more usefully describes relationship between salary and experience than model 1 with untransformed $Y$:
  - larger magnitude individual t-statistic.

• Note that $s$ and $R^2$ cannot be compared since the response variable is $\text{\$ thousands}$ for model 1 but $\log_e(\text{\$ thousands})$ for model 2.

• Model interpretation uses “multiplicative” idea.

• For example, since $\exp(\hat{b}_1) - 1 = \exp(0.059) - 1 = 0.0608$, we expect salary ($Y$) to increase by a multiplicative factor of 0.0608 (or 6.08%) for each additional year of experience ($X$).
• Investigate any relationship between annual taxes ($Y$) and sale price ($X$) for homes sold in Albuquerque, New Mexico in 1993.

• These data are typical of much business and economic data in that both variables are quite skewed (a few values much higher than the majority).

• What kind of transformation works well for (positively) skewed data?
• Investigate any relationship between annual taxes ($Y$) and sale price ($X$) for homes sold in Albuquerque, New Mexico in 1993.
• These data are typical of much business and economic data in that both variables are quite skewed (a few values much higher than the majority).
• What kind of transformation works well for (positively) skewed data?
• Which of these models might work better for this dataset?
  ○ Model 1: $E(Y) = b_0 + b_1 X$;
  ○ Model 2: $E(\log_e(Y)) = b_0 + b_1 \log_e(X)$. 
Plots include fitted regression (least squares) lines. Do both models satisfy the constant variance assumption?

Model 1 on the left: $E(Y) = b_0 + b_1X$.

Model 2 on the right: $E(\log(Y)) = b_0 + b_1\log(X)$.
### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Multiple R</th>
<th>R Squared</th>
<th>Adjusted R Squared</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.886&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.785</td>
<td>0.782</td>
<td>0.162</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Intercept), logX.

### Parameters<sup>a</sup>

| Model       | Estimate | Std. Error | t-stat | Pr(>|t|) |
|-------------|----------|------------|--------|---------|
| 2 (Intercept)| 2.076    | 0.237      | 8.762  | 0.000   |
| logX        | 0.983    | 0.051      | 19.276 | 0.000   |

<sup>a</sup> Response variable: logY.

So, $\log_e(\hat{Y}) = 2.076 + 0.983 \times \log_e(X)$

and $\hat{Y} = \exp(2.076 + 0.983 \times \log_e(X))$. 
Model 2 with $\log_e(Y)$ and $\log_e(X)$ more usefully describes relationship between annual taxes and sale prices than model 1 with untransformed $X$ and $Y$:

- model 2 satisfies regression assumptions, but model 1 fails constant variance assumption.

Model 2 is just as easy to use as model 1 for estimating or predicting annual taxes from home sale prices. For example, a home that sold for $100,000$ would be expected to have annual taxes of approx. $\exp(2.076 + 0.983 \times \log_e(100)) = 737$.

Prediction intervals should be more accurate for model 2 than for model 1 (which will tend to be too wide for low sale prices and too narrow for high sale prices)—see Problem 4.2.
Transformations in practice: example

4.1 Transformations
4.1.1 Natural logarithm transformation for predictors
4.1.2 Polynomial transformation for predictors
4.1.3 Reciprocal transformation for predictors
4.1.4 Natural logarithm transformation for the response
4.1.5 Transformations for the response and predictors

HOMETAX example
Scatterplots for models 1 and 2
Regression results for model 2
Interpretation

Transformations in practice: example

$E(\log_e(Y)) =$

$= b_0 + b_1 X_1 + b_2 X_2 + b_3 X_2^2 + b_4 \log_e(X_3) + b_5 (1/X_4)$.

- $Y$: natural logarithm transformation;
- $X_1$: untransformed;
- $X_2$: quadratic transformation (also included $X_2$ to retain hierarchy);
- $X_3$: natural logarithm transformation;
- $X_4$: reciprocal transformation.

© Iain Pardoe, 2006
4.1 Transformations

4.1.1 Natural logarithm transformation for predictors

4.1.2 Polynomial transformation for predictors

4.1.3 Reciprocal transformation for predictors

4.1.4 Natural logarithm transformation for the response

4.1.5 Transformations for the response and predictors

HOMETAX example
Scatterplots for models 1 and 2
Regression results for model 2
Interpretation

Transformations in practice: example

- \( E(\log_e(Y)) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_2^2 + b_4 \log_e(X_3) + b_5 (1/X_4). \)
  - \( Y \): natural logarithm transformation;
  - \( X_1 \): untransformed;
  - \( X_2 \): quadratic transformation (also included \( X_2^2 \) to retain hierarchy);
  - \( X_3 \): natural logarithm transformation;
  - \( X_4 \): reciprocal transformation.

- Best if transformations in a model are suggested before looking at the data from background knowledge about the situation or from theoretical arguments about why transformations make sense.

- Used judiciously, variable transformations provide a useful tool for improving regression models.

- Dangers: overcomplicating things unnecessarily, overfitting sample data (poor generalizability).