Stat 102

Introduction to Business Statistics
Class 8

Richard P. Waterman

Wharton

Table of contents I

Today's module

Topics to be covered in this module:

- Last time
- Not all relationships are linear
- Prospecting for curvature
- Data transformations
- The power function
- Comparing regressions on the original and transformed scale
- Summary
- Next time

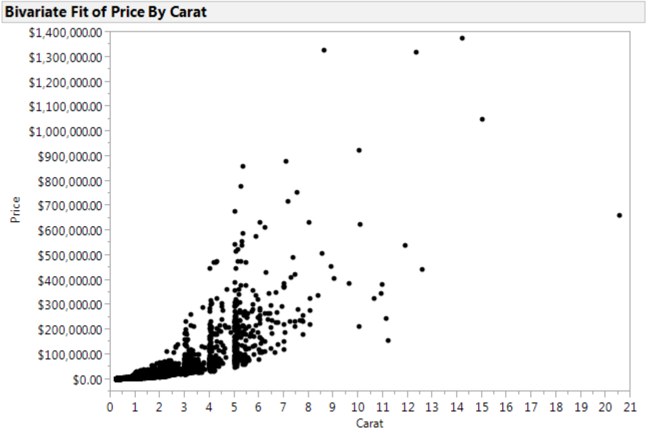
Last time

- The residual definition
- RMSE (s_e)
- Various residual plots
- R^2 and the quality of fit
- The relationship between R^2 and RMSE

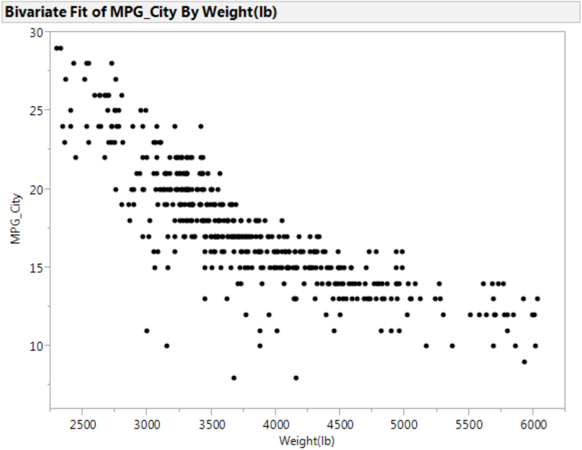
Some examples of non-linear patterns in real data

1. Diamond prices as a function of weight
2. Fuel economy as a function of weight
3. Quantity sold as a function of price
4. Quantity sold as a function of display feet
5. Number of people infected as a function of time
6. Number of cell phone subscribers as a function of time

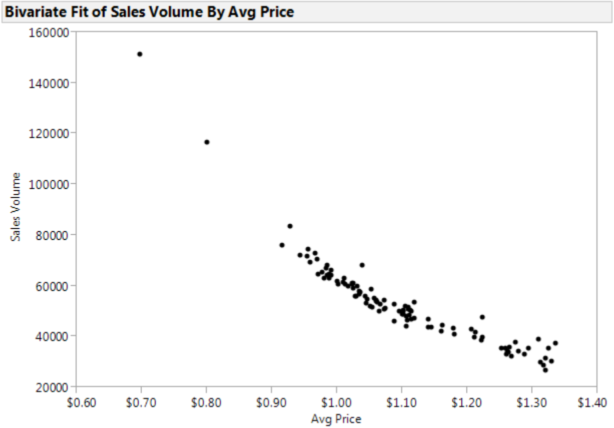
Diamond prices as a function of weight



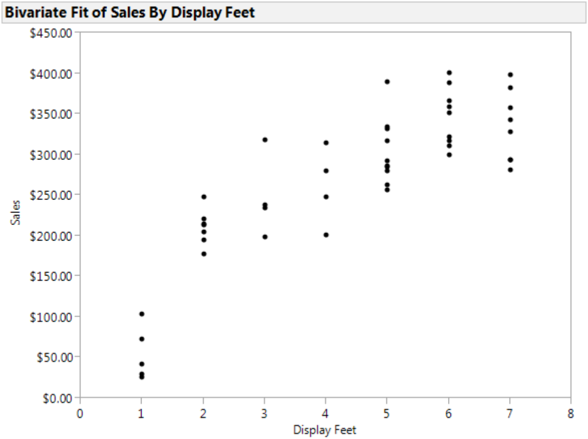
Fuel economy as a function of weight



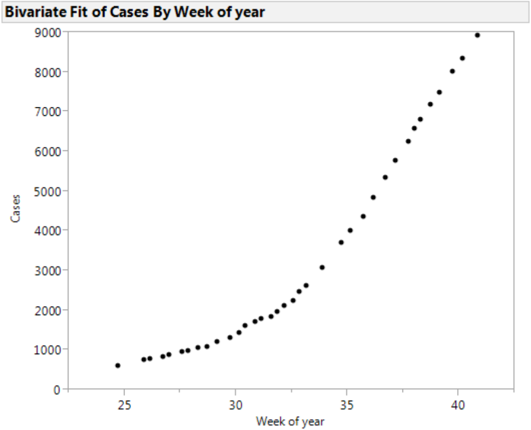
Quantity sold as a function of price



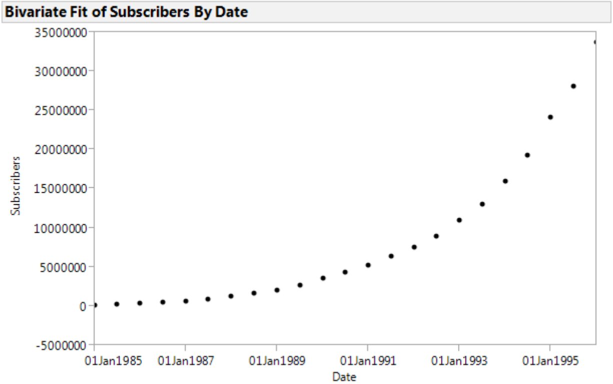
Quantity sold as a function of display feet



Number of people infected as a function of time



Number of cell phone subscribers as a function of time



Scatterplot smoothers

- A data driven way of prospecting for curvature
- Use the Flexible$\rightarrow$ Fit spline dialog from the scatterplot title bar
- Move the slider until you are happy with the fit
- You can become even *happier* by taking more Stat courses that will formalize choosing a bandwidth (moving the slider)

Choosing a transformation

Try some out:

- Think about the context
- Review the data
- Inspect the *scatterplot smooth*
- The rule in Stine p.515 gives some suggestions of transformations to try

The goal of the transformation

What makes a good transformation?

- We hope to see linearity on the transformed scale
- Fit the least squares line on the transformed scale
- For presentation purposes, back-transform the fit to the original scale of the data

Comments on the **power** function

The idea of a power function is to multiply a number by itself a given number of times. The number of times that the number has been multiplied by itself is called the *power*. The *power* is denoted with a superscript. It is sometimes called the *exponent* in the expression.

- n^1 is just n multiplied by itself once, and hence n itself.
- $n^2 = n \times n$ or n-squared.
- $n^3 = n \times n \times n$ or n-cubed.

Powers do not have to be whole numbers and they can also be negative. The two most frequently seen of these more exotic power functions are the square-root function (to the power one-half) and the reciprocal function (to the power minus one).

Defining the power function

We will now define the power function when the power is a whole number (integer) and then continue to use this definition to define negative and fractional powers.

Terminology: in the power function y^m, y is termed the **base** and m the *exponent*.

When m is an **integer** we define y^m as:

$$y^m = \underbrace{y \times y \times \cdots \times y}_{\text{m times}},$$

that is y multiplied by itself m times.

Example:

$$3^4 = \underbrace{3 \times 3 \times 3 \times 3}_{\text{4 times}} = 81.$$

The product of power functions

Using the definition of the power function, if we have two power functions with the same base multiplied together then:

$$y^m y^n = \underbrace{y \times y \times \cdots \times y}_{\text{m times}} \times \underbrace{y \times y \times \cdots \times y}_{\text{n times}} = \underbrace{y \times y \times \cdots \times y}_{\text{m + n times}} = y^{(m+n)}.$$

Example:

$$2^2 \times 2^3 = 2^{2+3} = 2^5 = 32.$$

This shows the key fact that the multiplication of power functions with the same base can be achieved by adding their exponents.

To the power 0

We can use the fact $y^m y^n = y^{(m+n)}$ to understand what happens when we raise a number to the power 0. By definition:

$$y^m \times y^0 = y^{m+0} = y^m.$$

This means that y^0 must be equal to 1 and gives rise to the Golden Rule: anything raised to the power 0 equals 1.

Power functions with a negative exponent

With this fact in hand we can consider what a negative exponent means by considering:

$$y^m \times y^{-m} = y^{m+(-m)} = y^0 = 1,$$

so that y^{-m} must equal $\frac{1}{y^m}$. In particular when $m = -1$ we have :

$$y^{-1} = \frac{1}{y}$$

often called the *reciprocal* function. This is one of the particular functions that is illustrated in the graph of power functions. The point to notice is that it models a function that is decreasing, that is one that goes from top left to bottom right in a graph.
In words: y to the negative m equals one over y to the power m.

Fractional exponents

So far we have only looked at exponents that are whole numbers, but exponents that are fractions are possible. We will start by looking at fractions that have one in the numerator, which are called the n-th roots, the most familiar of which is the square root function.
Start by raising a number to the power one half, then:

$$y^{\frac{1}{2}} \times y^{\frac{1}{2}} = y^{\frac{1}{2}+\frac{1}{2}} = y^1 = y,$$

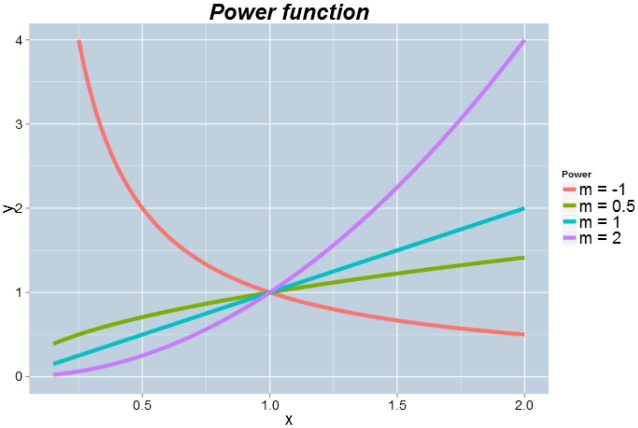
which means that $y^{\frac{1}{2}}$ is the number, which when multiplied by itself, gets you back to y again. This is the definition of the **square root**.
In general the function $y^{\frac{1}{m}}$ is called the m-th root of y. If you multiply this number by itself m times, you will get back y itself.
The m-th root function is written as:

$$\sqrt[m]{y}.$$

If you don't see an m in the root expression, then it is implicit that it is the square root of y.

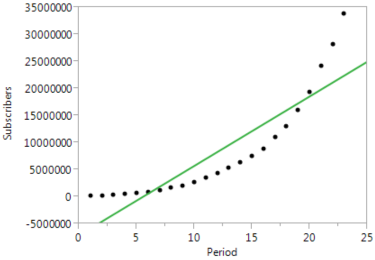
Examples of the **power** function

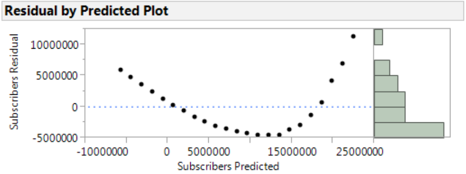
Figure 1: The graphs of some commonly-used power functions



Disaster.

Figure 2: A terrible "lack-of-fit"

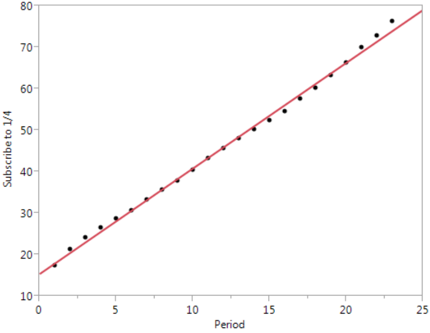




The cell phone data set

We will use the transformation $y^{\frac{1}{4}}$.

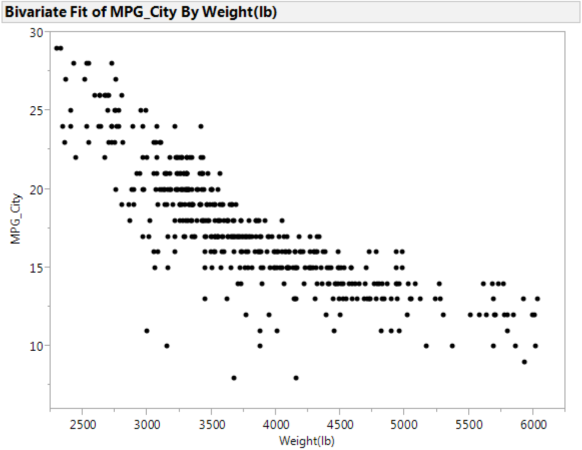
Figure 3: A much better linear fit



Making a prediction

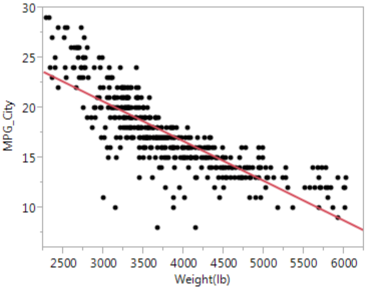
- What does the model predict at t = 24?
- $$\hat{y}^{\frac{1}{4}} = 15.377 + 2.5474 * Period.$$
- $$\hat{y}^{\frac{1}{4}} = 76.5146.$$
- Back transform with the "to the power four" function to the original scale of the data:
$$76.5146^4 = 34,274,983$$

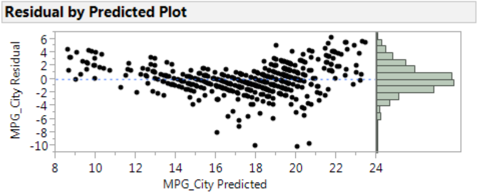
Fuel economy as a function of weight



Linear fit

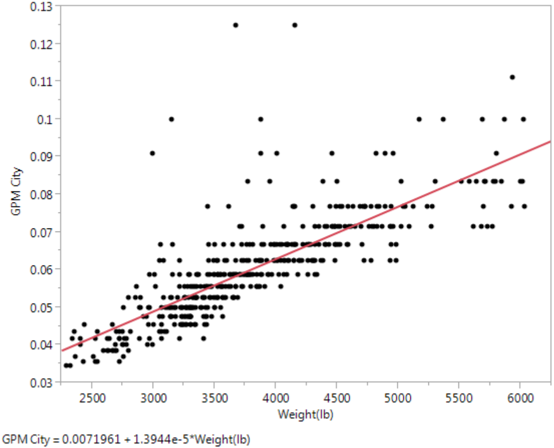
Once again, not a good fit at all.





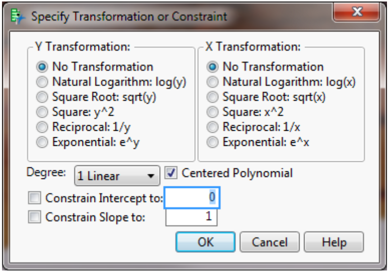
The $1/y$ transform (y^{-1})

Not perfect, but much better



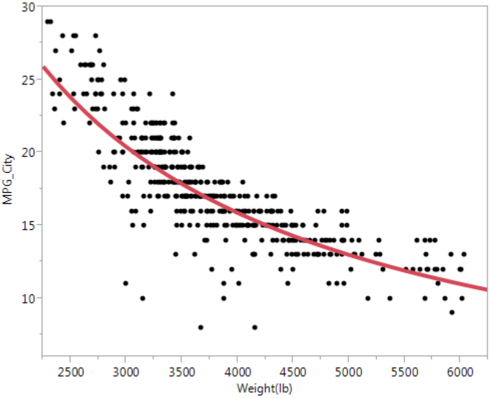
The JMP transformation dialog

Using "Fit special" from the scatterplot makes things very quick.



The $1/y$ transform

The same model as before but now plotted back on the original scale



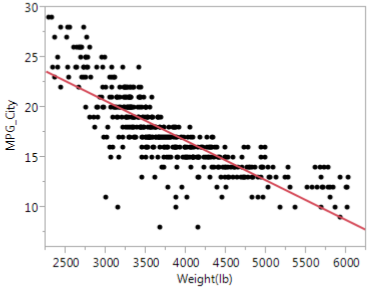
Comparing models on the original and transformed scales

- A direct comparison using the two main model summaries, RMSE and R^2 is problematic if Y is on different scales.
- RMSE has the scale of Y, so direct comparisons across different measurement scales is meaningless.
- R^2 measures the proportion of variability in Y explained by the regression model.
- Knowing how much variability is explained on a transformed scale doesn't tell you how much variability is explained on the original scale of the data. For example, you get paid in dollars, not log dollars or square root dollars.
- The solution: back-transform the fit on the transformed scale and use this to recalculate RMSE and R^2 on the original scale. Now you can compare like with like.

Model comparisons

1. The original bad linear fit
2. Fit on transformed scale
3. Back-transformed fit, now on the original scale

1. The original bad linear fit

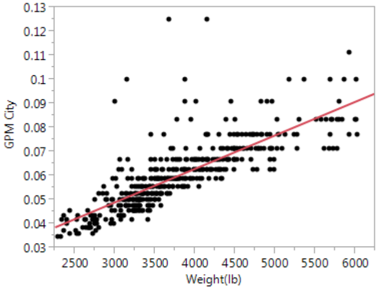


Summary of Fit

RSquare	0.669933
RSquare Adj	0.669243
Root Mean Square Error	2.219958
Mean of Response	17.39792
Observations (or Sum Wgts)	480

Note R^2 and RMSE.

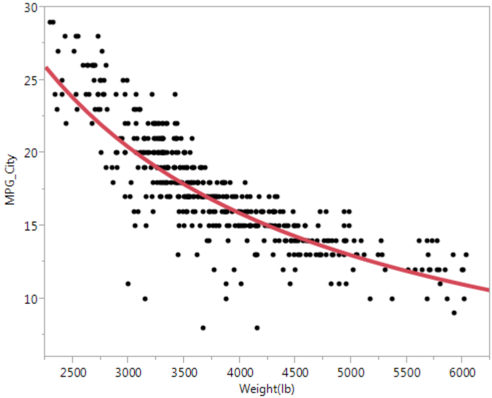
2. Fit on transformed scale



Summary of Fit	
RSquare	0.638682
RSquare Adj	0.637926
Root Mean Square Error	0.008361
Mean of Response	0.0604
Observations (or Sum Wgts)	480

Note R^2 and RMSE. These are not comparable to the previous summaries because they are on the transformed scale.

3. Back-transformed fit, now on the original scale



Fit Measured on Original Scale	
Sum of Squared Error	1919.9243
Root Mean Square Error	2.0041402
RSquare	0.7309899
Sum of Residuals	146.84067

Note R^2 and RMSE for the *fit measured on the original scale*. These are comparable to the original fit, because they are calculated using Y back on the original scale.

Valid comparisons

- Comparisons between 1 and 3 are valid because the outcome is on the same scale.
- 2 is on a different scale (1/y) so not directly comparable to 1.
- Notice that from 3 v. 1, RMSE is lower (2.00 v. 2.22) and R^2 is higher (73% v. 67%).
- This shows that the transformation has been successful.
- So long as you use the "fit model" dialog, JMP will summarize the fit from the transformation back on the original scale. This is very useful.

Module summary

1. Prospecting for curvature
2. Data transformations
3. The power function
4. Comparing regressions on the original and transformed scale

Next time

- Log transforms and optimization