Wharton
UNIVERSITY of PENNSYLVANIA

Nonlinear Models… Fitting Curves

ROADMAP

Linear relation: effect on y of changes in x is the same at every value of x.

Nonlinear relation: effect on y of changes in x depends on the value of x.

We can expect nonlinearities in many business applications
- Diminishing marginal effect (eg, promotion response, manufacturing)
- Relationships with constant elasticity (eg, price and demand)

Examples

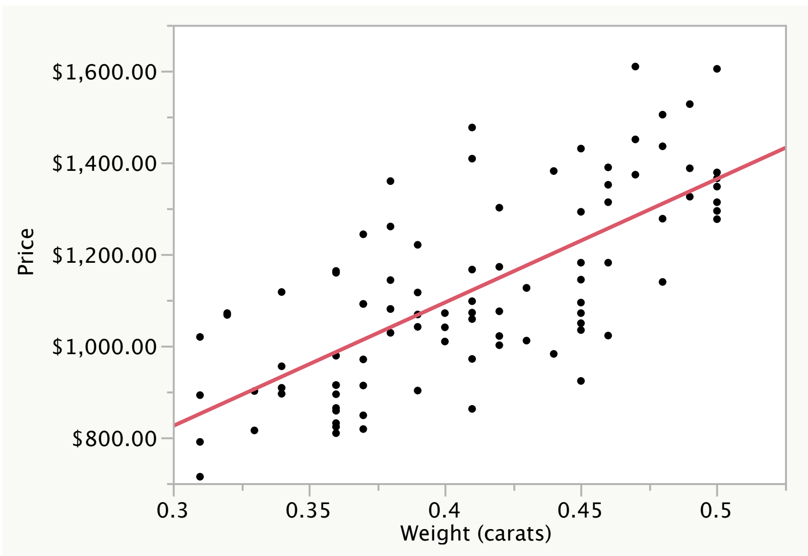
Diamond prices	diamonds.jmp, more_diamonds.jmp
Track times	track.jmp
Retail sales	display.jmp
Auto mileage	cars.jmp

LINEAR MODELS

Meaning of linearity

Equal changes in X associated with equal changes in Y (on average)

Example: diamond prices



Linear Fit

Price = 15.199987 + 2697.2532*Weight (carats)

Summary of Fit

RSquare	0.508585
RSquare Adj	0.503184
Root Mean Square Error	144.8662
Mean of Response	1119.624
Observations (or Sum Wgts)	93

Lack Of Fit

Analysis of Variance

Parameter Estimates

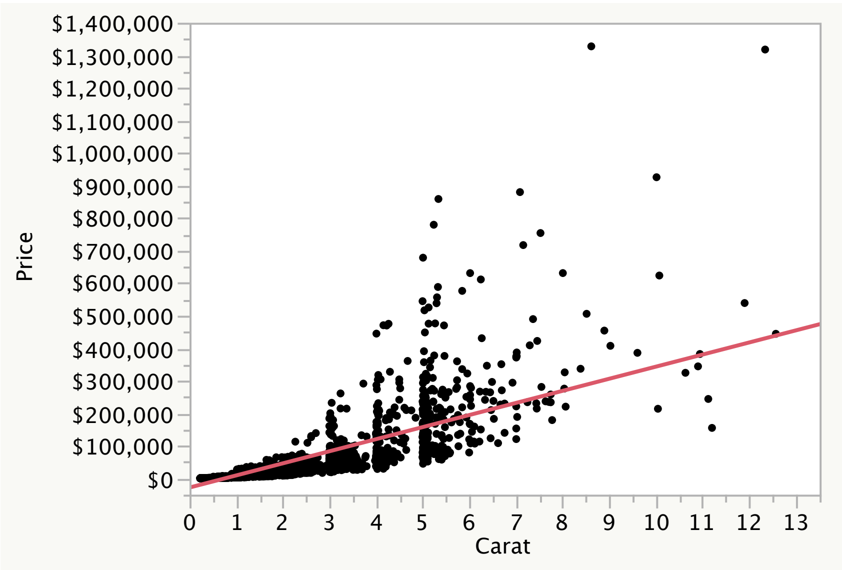
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15.199987	114.7912	0.13	0.8949
Weight (carats)	2697.2532	277.9354	9.70	<.0001*

Interpretation:

Model parameters, assumptions

Substantively: Does this model make sense?

Add more data, namely larger diamonds



Linear Fit

Price = -28047.8 + 37116.168*Carat

Summary of Fit

RSquare	0.587134
RSquare Adj	0.58708
Root Mean Square Error	41519.61
Mean of Response	21070.99
Observations (or Sum Wgts)	7568

Lack Of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-28047.8	672.3232	-41.72	<.0001*
Carat	37116.168	357.8208	103.73	<.0001*

Does a linear model still make sense?

How do these data violate conditions implied by the SRM?

Could we have anticipated these problems earlier?

What to do about these problems?

NONLINEAR MODELS… CURVES

Question: What sort of curve captures the pattern in the prior plot?

Finding the right transformation…

(a) Graphically from scatterplot of Y on X

(b) Residual plots

(c) Substantively (what would make sense)

Logs and percentages

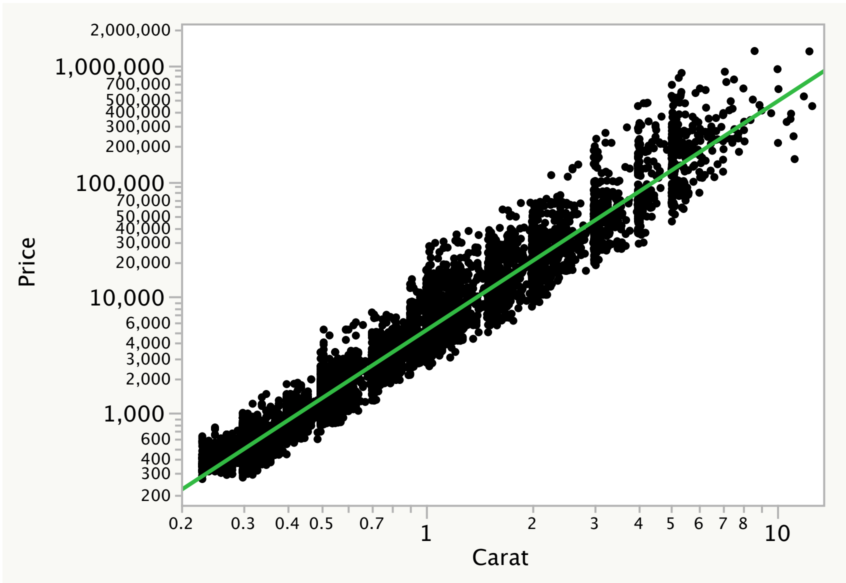
Change on a log scale: think percentages

$$
\begin{aligned}
\log_e(x) - \log_e(y) &= \log_e(x/y) \\
&= \log_e\left((y + x\text{-}y)/y \right) \\
&= \log_e\left(1 + (x\text{-}y)/y\right) \\
&\approx (x\text{-}y)/y \qquad \text{if } x\text{-}y \text{ is small compared to } y^{1}
\end{aligned}
$$

Modeling prices: How are diamond prices related to the price of diamonds?

[1] This only works in such a nice way with natural logs. Base 10 logs (or others) require some messy constants.

On log scales, the fit appears linear…[2]



Transformed Fit Log to Log

Log(Price) = 8.5540316 + 1.972383*Log(Carat)

Summary of Fit

RSquare	0.945776
RSquare Adj	0.945769
Root Mean Square Error	0.387741
Mean of Response	8.398948
Observations (or Sum Wgts)	7568

Lack Of Fit

Analysis of Variance

Parameter Estimates

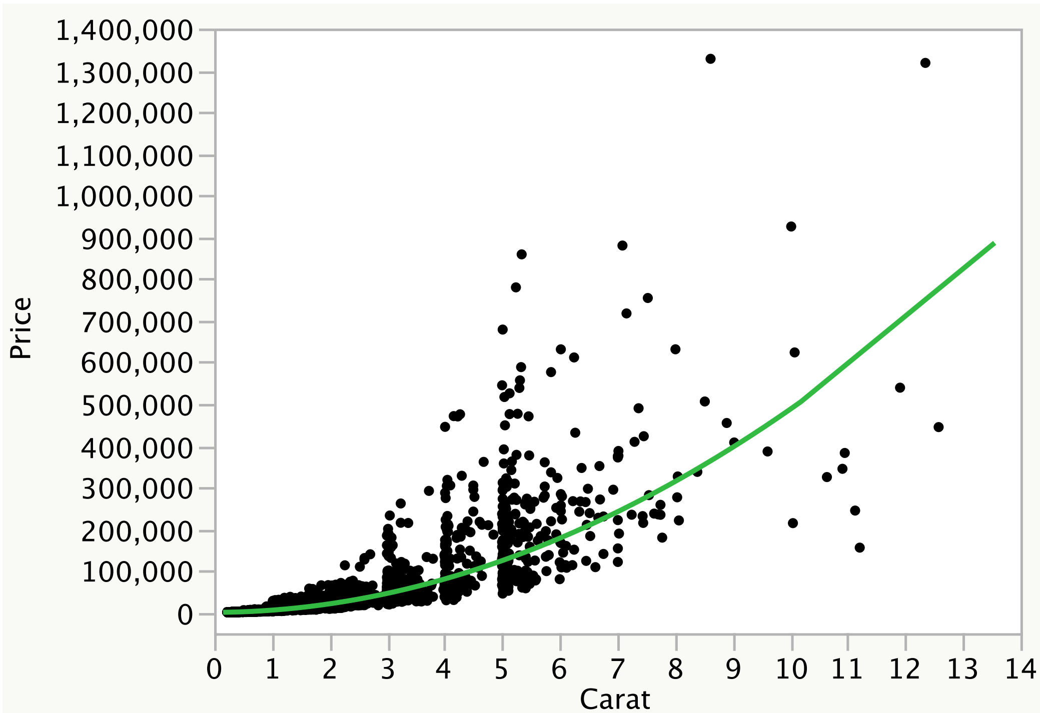
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.5540316	0.004477	1910.5	<.0001*
Log(Carat)	1.972383	0.00543	363.27	<.0001*

Interpretation of slope: elasticity of price with respect to size in carats.

1% increase in weight associated with 2% (1.97) increase in price, on avg

[2] Do this in JMP by double clicking on each axis in the scatterplot and picking the "log" option in the dialog. To get the fitted model using logs, select the Fit special item from the Fit Y by X dialog. An example of this dialog appears later.

Fit on original scale…



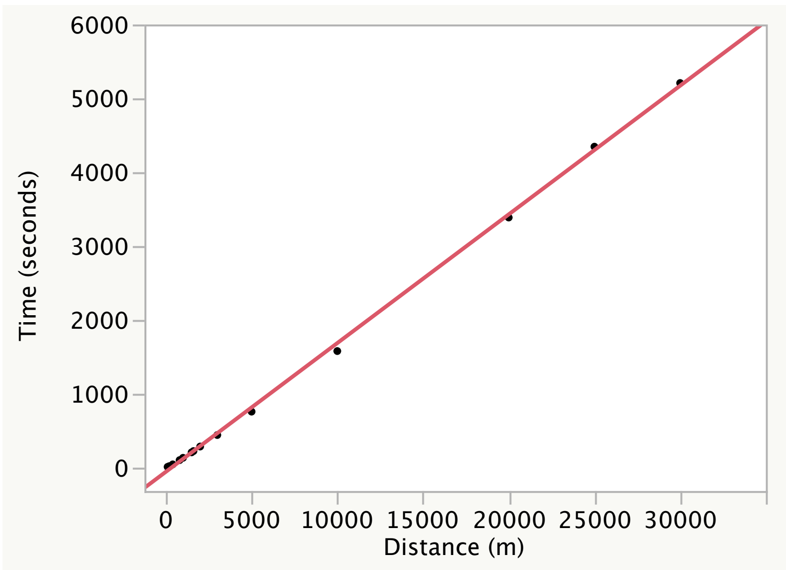
"Linear regression"…

The fitted model always has a "linear" equation, but the variables X and Y may involve transformations of the original measurements.

$$\underbrace{\text{Estimated } (\log_e \text{Price})}_{Y} = 8.554 + 1.972\ \underbrace{(\log_e \text{Carat})}_{X}$$

VALUE OF RESIDUAL PLOTS

Nonlinear patterns are not always visible in the scatterplot itself, and only become apparent in the detail offered by the residual plot. Consider the following men's records in track events



Linear Fit

Time (seconds) = −57.1417 + 0.1742686*Distance (m)

Summary of Fit

RSquare	0.999354
RSquare Adj	0.9993
Root Mean Square Error	46.94126
Mean of Response	1195.219
Observations (or Sum Wgts)	14

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	−57.1417	15.55293	−3.67	0.0032*
Distance (m)	0.1742686	0.001279	136.24	<.0001*

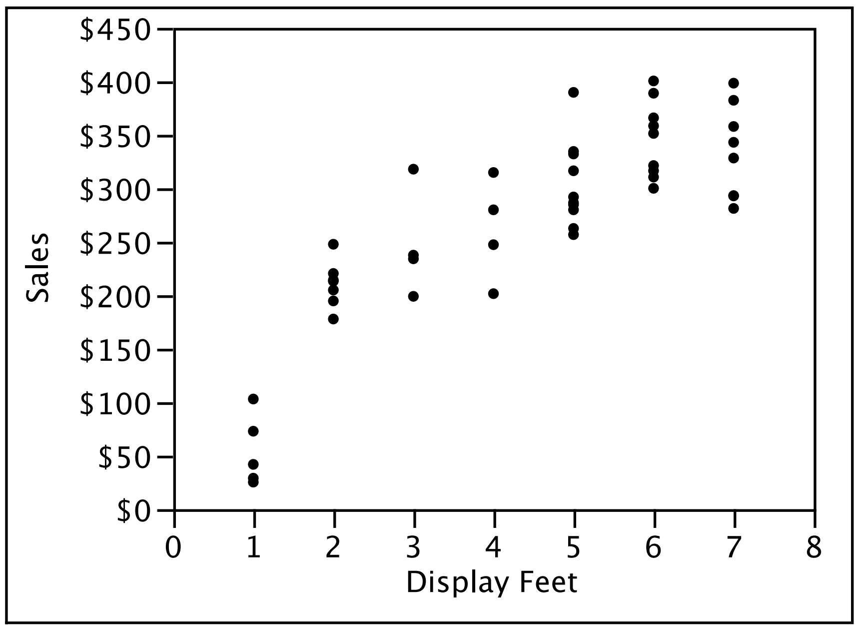
With R^2 so large, can there possibly be a problem with this model?

ANOTHER NONLINEAR REGRESSION

Common question: what is the optimal amount of promotion for a product?

Specific case: A chain of liquor stores needs to know how much shelf space in its stores to devote to showing a new wine to maximize its profit. Space devoted to other products brings in about $50 of net revenue per linear foot.

The data *Display.jmp* has weekly sales ($) and shelf-feet from 47 stores of the chain. Should we expect a linear relationship between promotion and sales, or should we expect diminishing marginal gains?



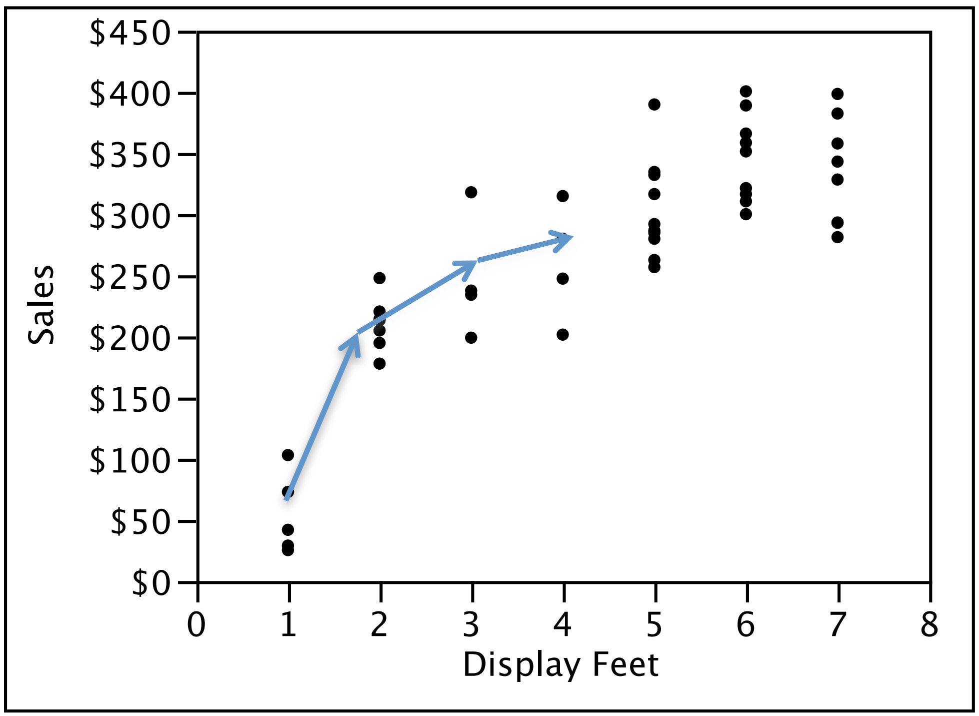
Visually, what do the data suggest for the optimal shelf space?

DETERMINING THE OPTIMAL AMOUNT

Key economic principle

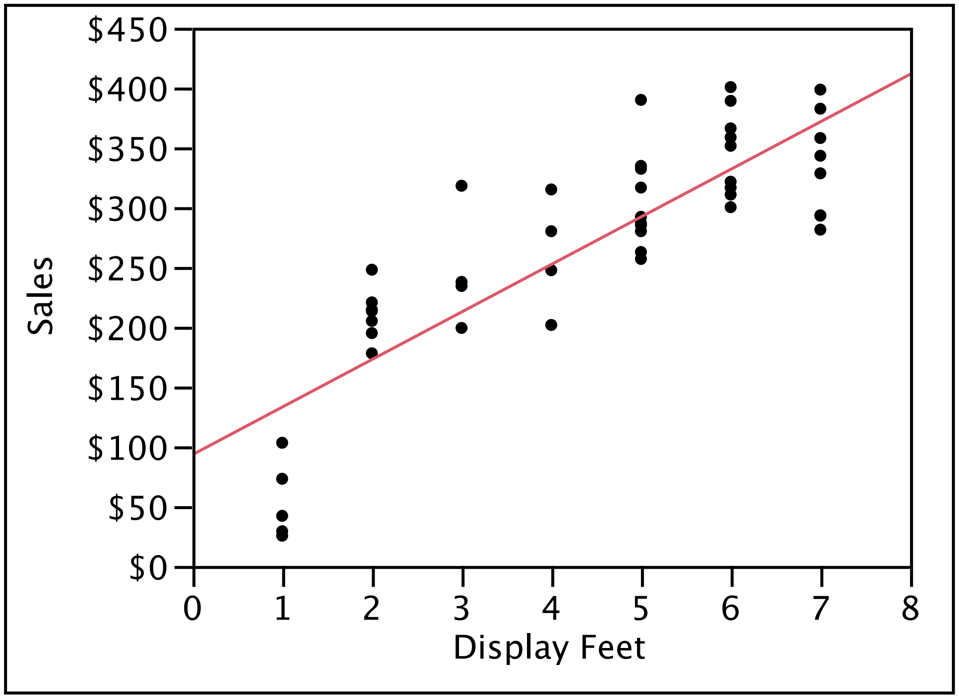
Find the amount of space at which Marginal cost = Marginal revenue

From the graph, imagine connecting the averages in each column of points with a line and inspect how sales changes.



Using a line to predict sales (y) for a given amount of promotion (x) seems silly. If sales constantly increased with amount of promotion, that would imply we should either (a) sell nothing but this product or (b) not sell any of this product.

Also, the LS regression line misses some of the different group means, particularly when little material is on display.



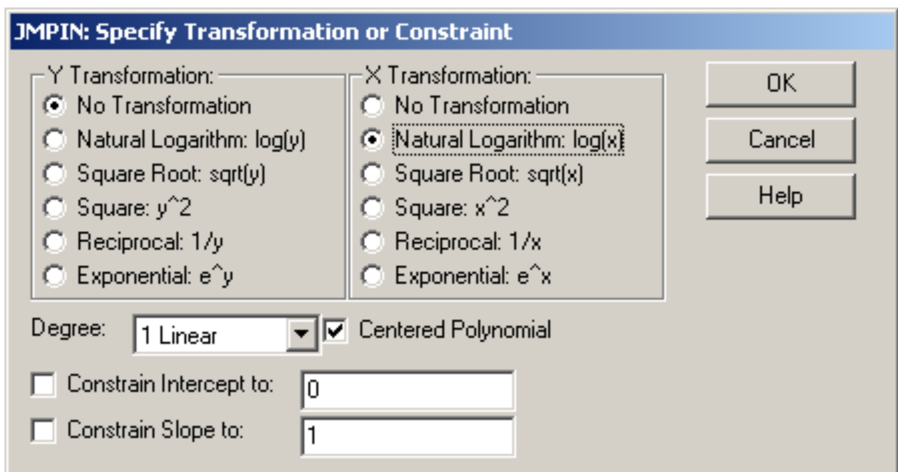
Est Sales
= $93 + 39.76 Ft

To get a feel for the shape of the relationship between sales and display feet, you could simply sketch a smooth curve that is closer to the center of each group.

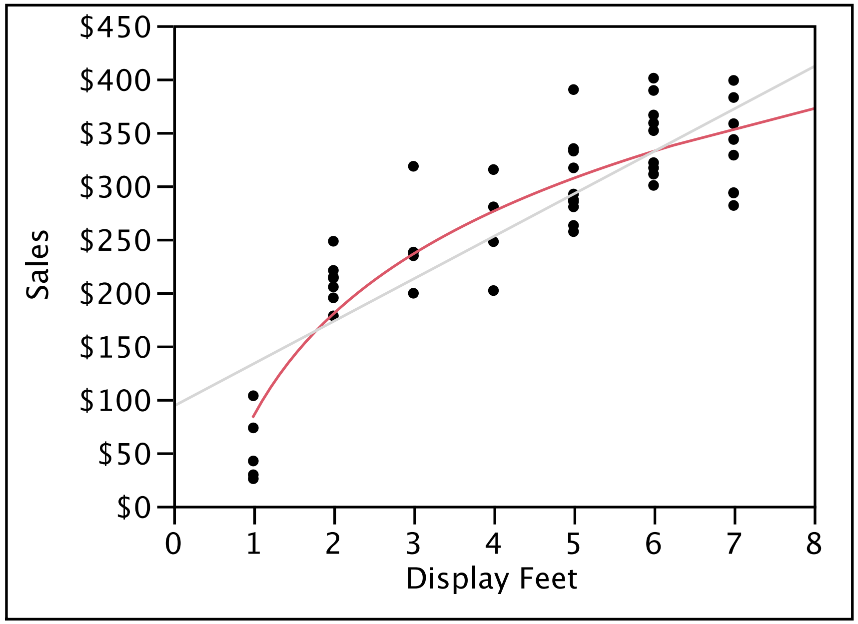
The shape of this curve is similar to the shape of $y = \log_e x$, and so we might consider fitting a curve of the form

$$y = b_0 + b_1 \log_e x$$

This can be done in JMP using the Fit Y by X subcommand Fit Special and selecting Natural Logarithm for the X transformation.



This command fits the equation of a curve that bends as shown on the next page.



The fitted equation

$$y = 83.56 + 138.62 \log_e x$$

is still a "least squares fit" in the sense that of all functions $f(x) = b_0 + b_1 \log_e x$, this one minimizes the sum of squared vertical deviations from the equation to the data.

Visually, the fitted equation does a reasonable job describing the relationship between average sales and display feet. In order to decide whether this equation makes sense, we need to interpret the equation in the context of the problem.

Optimal shelf space. Combine this equation with the fact that each shelf foot otherwise provides \$50 of net revenue. How much shelf space should be devoted to the new wine?

Solution: Find the point where marginal cost = marginal revenue. That is, find the number of feet where the slope is \$50. Since the derivative of $f(x)$ is $f'(x)=b_1/x$, the optimal amount is $x_{opt} = 138.62/50 \approx 2.8$ feet[3].

Interpretation of slope and intercept. Recall that variation on the scale of natural logs is comparable to variation in percentages.

What is the interpretation of $b_1 = 138.62$?

> If x were 1% larger, then y would be about \$1.4 higher.[4]

What is the interpretation of $b_0 = 83.56$?

> $y = 83.56$ when x is chosen so that $\log_e(x) = 0$ which is when $x = 1$.

How could you add confidence intervals to these, particularly the optimal space?

---

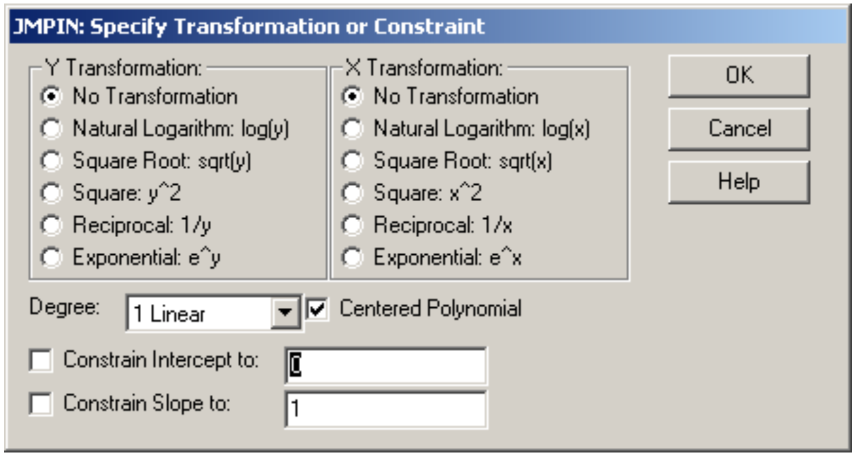
[3] How precise do we need this estimate? What is the natural "granularity of the problem"? Ans.: the width of a bottle of wine.

[4] Use the fact that $\log_e(1.01x) = \log_e(1.01) + \log_e x \approx .01 + \log x$ for log base e. To convince yourself of this approximation, use your calculator to compute $\log_e(1.01)$.

OTHER TRANSFORMATIONS

The previous choice of a transformation for the non-linear regressions was guided by the shape of the relationships and the sensibility of its interpretation. However, other choices might also be reasonable.

The subcommand Fit Special offers a variety of such choices for transforming *y* and/or *x*.



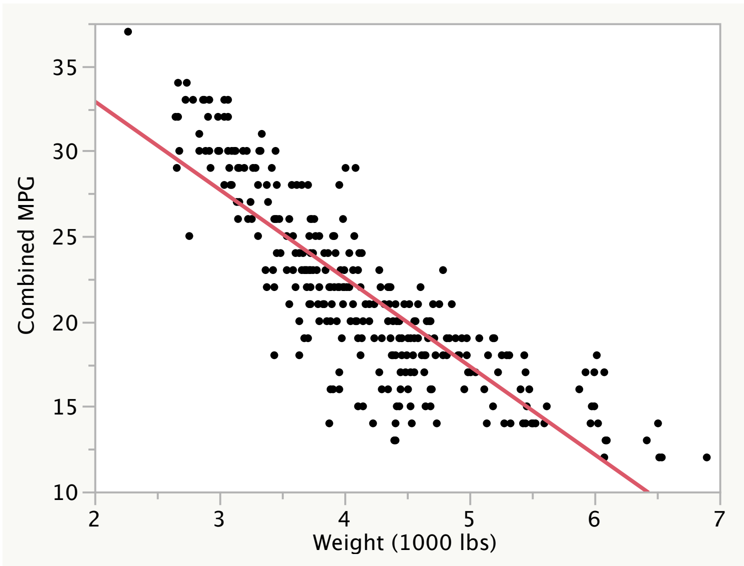
One typically chooses the best transformation by combining what you know about the data and application with an exploratory process.[5]

Logs are most common in business applications.

[5] Have a look at Figure 20.5 in SF (page 515) for hints on picking a useful transformation. Unless they fit poorly, we generally use logs, and occasionally reciprocals. These produce interpretable equations.

EXAMPLE OF OTHER TRANSFORMATIONS

Relationship between fuel consumption and vehicle weight…



Linear Fit

Combined MPG = 43.299846 – 5.1857491*Weight (1000 lbs)

Summary of Fit

RSquare	0.702469
RSquare Adj	0.70148
Root Mean Square Error	2.946822
Mean of Response	21.66997
Observations (or Sum Wgts)	303

Lack Of Fit

Analysis of Variance

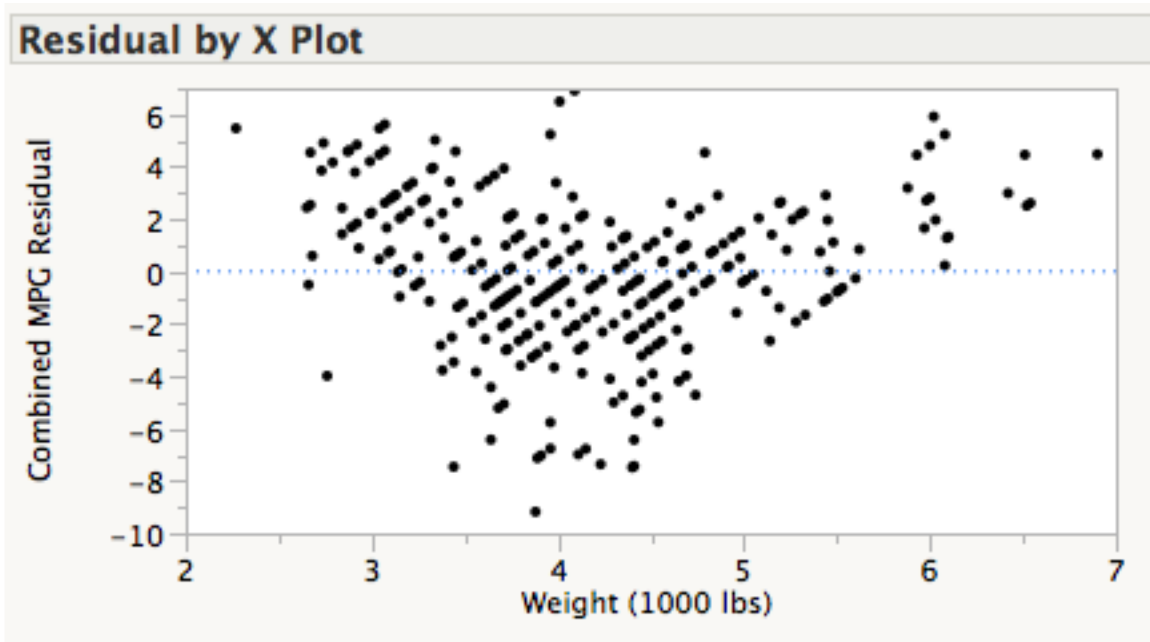
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	43.299846	0.828852	52.24	<.0001*
Weight (1000 lbs)	–5.185749	0.194528	–26.66	<.0001*

It's an okay fit (the line captures the downward pattern), but does it make sense?

BTW, which vehicles are these?

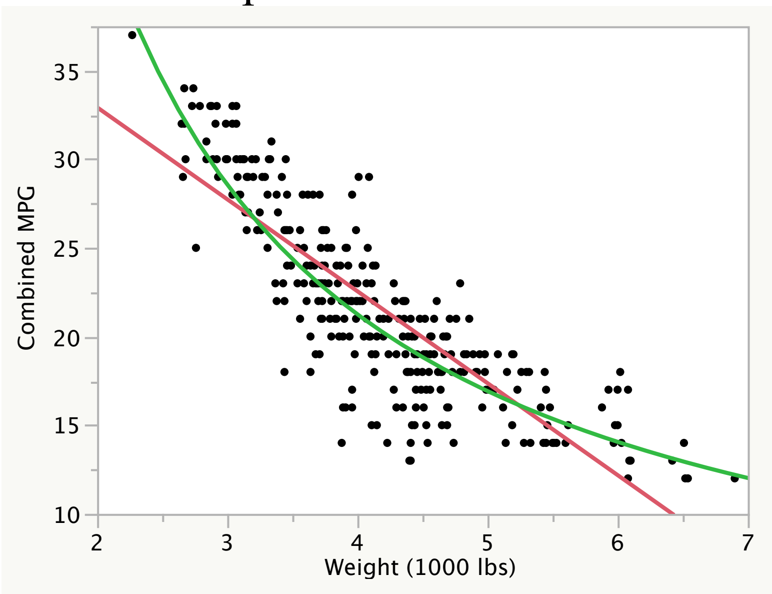
Residual plots make the problem easier to see.



The fit (dotted blue line) tends to be too low (underpredicts) for light and heavy vehicles, and too high for medium weight vehicles.

A variety of transformations produce a more sensible model. For this example, the reciprocal produces an equation that is familiar (at least in Europe).

Transforming the response from MPG to 1/MPG captures the curvature and has a more natural extrapolation.



Transformed Fit Reciprocal

Recip(Combined MPG) = -0.00112 + 0.0120418*Weight (1000 lbs)

Summary of Fit

RSquare	0.713207
RSquare Adj	0.712254
Root Mean Square Error	0.006667
Mean of Response	0.049106
Observations (or Sum Wgts)	303

Lack Of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.00112	0.001875	-0.60	0.5507
Weight (1000 lbs)	0.0120418	0.00044	27.36	<.0001*

Interpretation:

How does the reciprocal scale capture the effect of reducing the weight of a vehicle?

Does reducing the weight of a big truck by 200 lbs have the same effect on the miles per gallon as reducing the weight of a small compact car?

Going further: What problem happens with the reciprocal transformation?[6]

[6] To answer this, use a formula to compute the reciprocal of MPG and plot the reciprocal on the weight. Do you see a problem?

LECTURE REVIEW

A regression line offers a summary of the relationship between a predictor (called x) and a response (called y).

Transformations of variables (most often logarithms) allow regression to capture nonlinear patterns as well.

The interpretations of the slope and intercept depend on the specific transformation. Logs are associated with percentage changes.

The slope in a log-log model gives the elasticity, associating constant percentage changes in x with constant percentage changes in y.

APPENDIX: LOGS IN REGRESSION

The interpretation in percentages assumes that these are base-e or natural logs.

y on x Linear As x increases by 1, avg y increases by b_1	y on $\log(x)$ Diminishing returns As x increases by 1%, avg y increases by 0.01 b_1
$\log(y)$ on x Exponential growth As x increases by 1, avg y increases by 100 b_1%	$\log(y)$ on $\log(x)$ Demand curve As x increases by 1%, avg y increases by b_1%