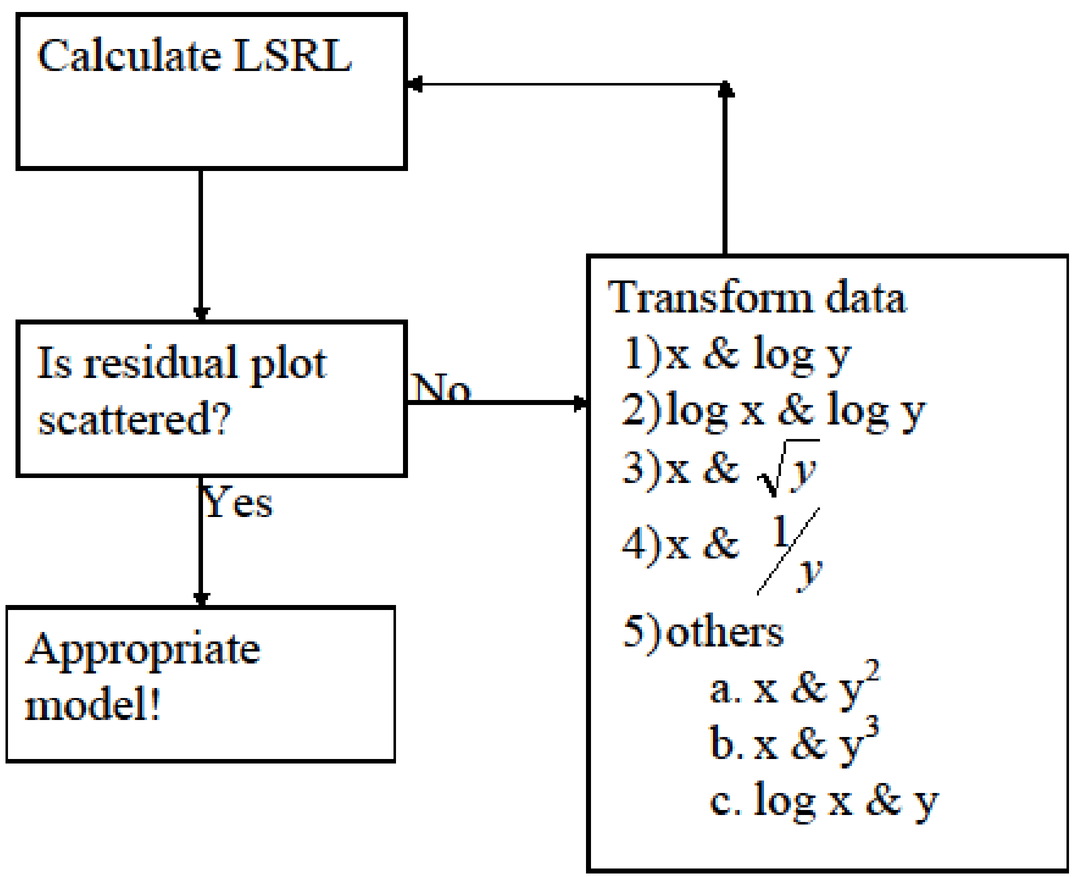
5.4 Transforming to Achieve Linearity

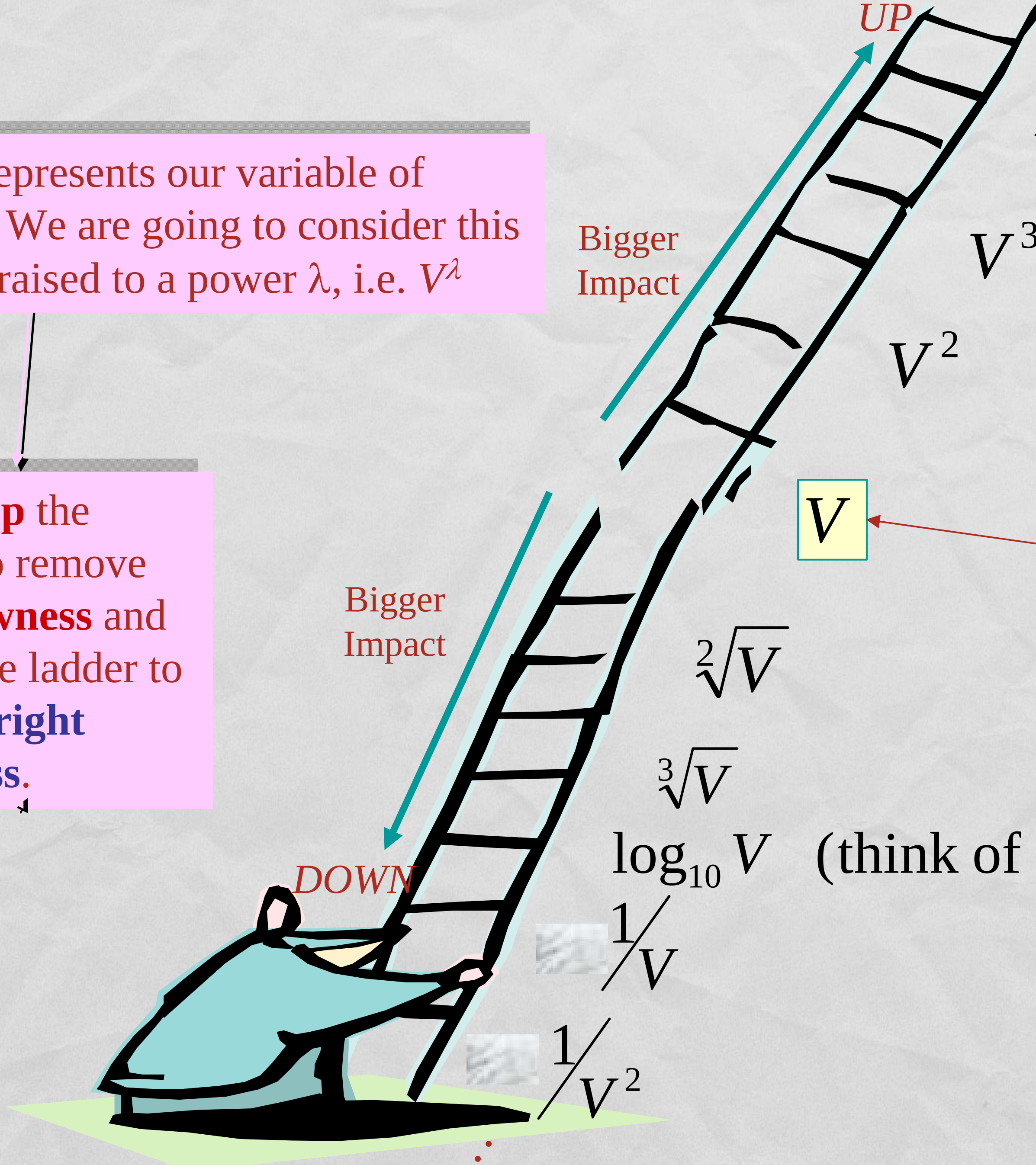
What to do if data is non-linear:



Ladder of Powers

Here V represents our variable of interest. We are going to consider this variable raised to a power λ, i.e. V^{λ}

We go **up** the ladder to remove **left skewness** and **down** the ladder to remove **right skewness**.



$$V^4$$

$$V^3$$

$$V^2$$

$$V$$

$$\sqrt[2]{V}$$

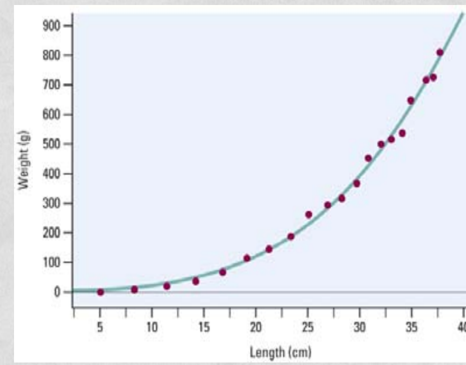
$$\sqrt[3]{V}$$

$$\log_{10} V \quad (\text{think of this as } V^0)$$

$$1/V$$

$$1/V^2$$

Middle rung:
No transformation
(λ = 1)
Linear Model is appropriate



Watch Video

David Bock “Ladder of Powers”

- http://media.pearsoncmg.com/cmg/pmmg_mml_shared/flash_video_player/player.html?aw/aw_deveaux_introstats_3/video/stat3dv_1000

Transforming with Powers (don't memorize – see examples next slide)

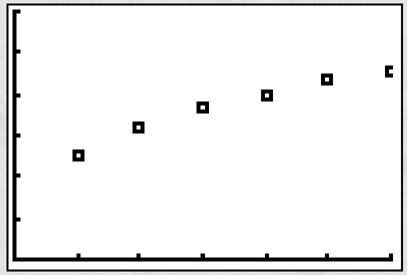
- Facts about powers:
  - The graph of a power with exponent 1 (p = 1) is a straight line.
    - Powers greater than 1 give graphs that bend upward. The sharpness of the bend increases as the power increases.
    - Powers less than 1 but greater than 0 give graphs that bend downward.
    - Powers less than 0 give graphs that decrease as x increases. Greater negative values of p result in graphs that decrease more quickly.
    - The logarithm function corresponds to p = 0 (not the same as raising to the 0 power which is just a horizontal line at y = 1)

Here are Samples of Graphs

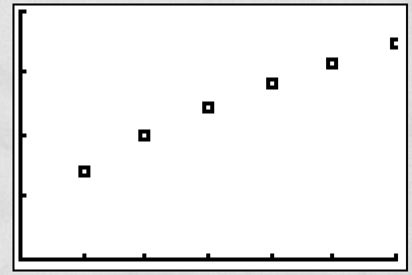
and

the Transformations to create a linear association

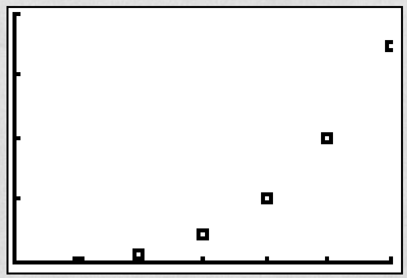
Transformation x and y^3



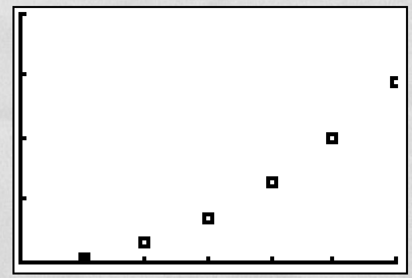
Transformation x and y^2



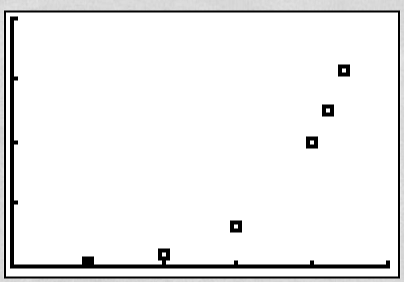
Transformation x and y^(1/3)



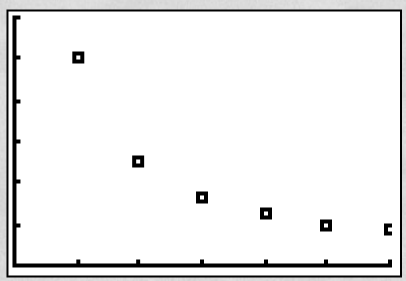
Transformation x and y^(1/2)



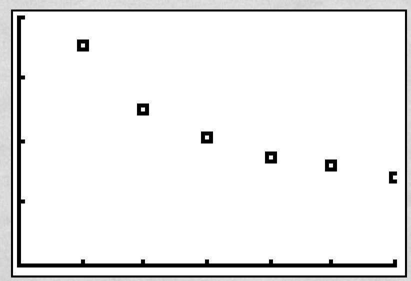
Transformation x and log(y)



Transformation x and 1/y



Transformation x and $1/y^2$



The Logarithm Transformation

- If an exponential model of the form $y = ab^x$ describes the relationship between x and y then we can use logarithms to transform the data to produce a linear relationship (and vice versa- if a transformation of (x,y) data to (x, log y) straightens our data, we know it's exponential

Algebraic Properties of Logarithms

$$\log_b x = y \quad \text{if and only if} \quad b^y = x$$

The rules for logarithms are

1. $\log_b (MN) = \log_b M + \log_b N$
2. $\log_b (M/N) = \log_b M - \log_b N$
3. $\log_b X^p = p\log_b X$

- So how does this work? well if we have the equation $y = ab^x$ and take the log of both sides:
  - $\log y = \log (ab^x)$
    - $= \log a + \log b^x$
    - $= \log a + \log b\,(x)$ Does this look familiar?!

Summary of what you need to know about log transformations

- When data doesn't look straight, try both transformations: (x,y) to (x, logy) or (x, lny) and (logx, logy) or (lnx, lny)- **log and natural log are both fine!**
- Check which transformation did a better job straightening:
  - Make a scatterplot of each transformation. Do LinReg a+bx to check your r for each. The stronger the r, the better.
  - Also do a residual plot for each transformation to see which better fits the data (for exponential trial: L1, RESID. For Power Law trial: L3, RESID)
  - If your first transformation was better than it's an underlying exponential function fitting your data. If the second transformation was better than it's a power model.
  - Find the regression equation for your original untransformed data:
    - If it was exponential, yhat = (10^a)(10^b)^x
    - If it was a power model, yhat = (10^a)(x^b)