Regression III: Advanced Methods

William G. Jacoby
Michigan State University

Goals of the lecture

- The Ladder of Roots and Powers
- Changing the shape of distributions
- Transforming for Linearity

Why transform data?

1. In some instances it can help us better examine a distribution
2. Many statistical models are based on the mean and thus require that the mean is an appropriate measure of central tendency (*i.e.*, the distribution is approximately normal)
3. Linear least squares regression assumes that the relationship between two variables is linear. Often we can "straighten" a nonlinear relationship by transforming one or both of the variables
    - Often transformations will 'fix' problem distributions so that we can use least-squares regression
    - When transformations fail to remedy these problems, another option is to use nonparametric regression, which makes fewer assumptions about the data

Power transformations for quantitative variables

- Although there are an infinite number of functions *f(x)* that can be used to transform a distribution, in practice only a relatively small number are regularly used
- For quantitative variables one can usually rely on the "family" of powers and roots:

$$X \rightarrow X^p$$

- When *p* is negative, the transformation is an inverse power:

$$X^{-1} = 1/X,\ X^{-2} = 1/X^2,\ X^{-3} = 1/X^3, etc$$

- When *p* is a fraction, the transformation represents a root:

$$X^{1/2} = \sqrt{X},\ X^{-1/2} = 1/\sqrt{X}, etc.$$

Log transformations

- A power transformation of X^0 should not be used because it changes all values to 1 (in other words, it makes the variable a constant)
- Instead we can think of X^0 as a shorthand for the log transformation $\log_e X$, where e * 2.718 is the base of the natural logarithms:

$$\lim_{p \to 0} \frac{X^p - 1}{p} = \log_e X$$

- In practice most people prefer to use $\log_{10} X$ because it is easier to interpret—increasing $\log_{10} X$ by 1 is the same as multiplying X by 10
- In terms of result, it matters little which base is used because changing base is equivalent to multiplying X by a constant

Cautions: Power Transformations (1)

- Descending the ladder of powers and roots compresses the *large values* of X and spreads out the *small values*
- As *p* moves away from 1 in either direction, the transformation becomes more powerful
- Power transformations are sensible ONLY when all the X values are POSITIVE—If not, this can be solved by adding a start value
  - Some transformations (e.g., log, square root, are undefined for 0 and negative numbers)
  - Other power transformations will not be monotone, thus changing the order of the data

X	X^2
-2	4
0	0
1	1
2	4

X	$(X+3)^2$
-2	1
0	9
1	16
2	25

Cautions: Power Transformations (2)

- Power transformations are only effective if the ratio of the largest data value to the smallest data value is large
- If the ratio is very close to 1, the transformation will have little effect
- *General rule*: If the ratio is less than 5, a negative start value should be considered

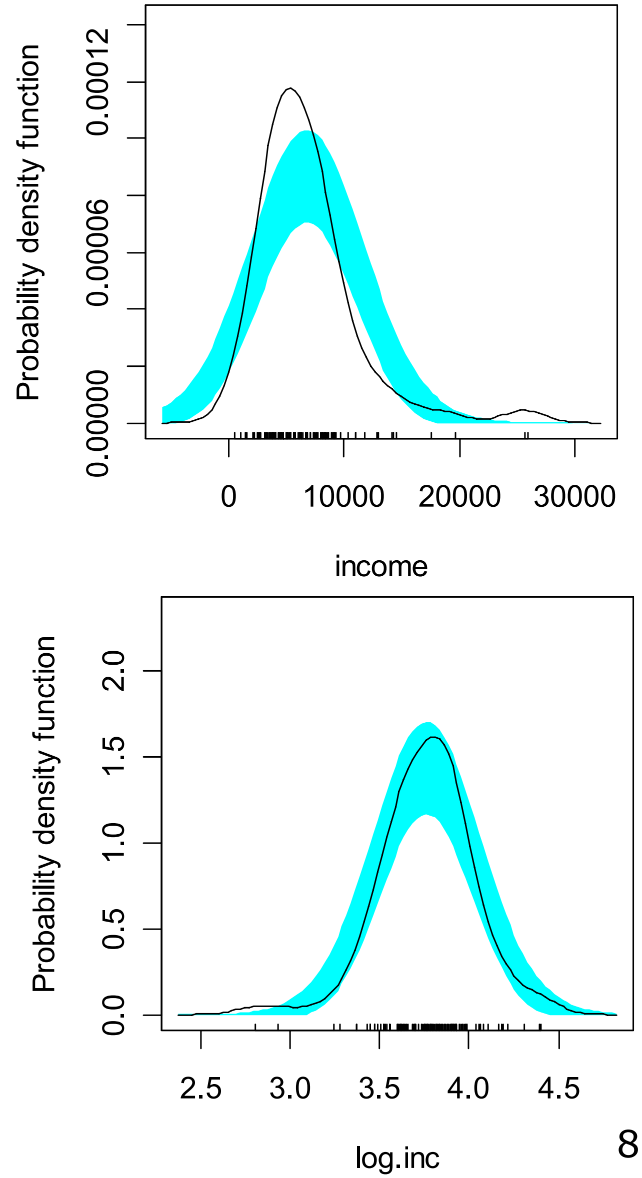
X	$\log_{10} X$
1991	3.2990
1992	3.2992
1993	3.2995
1994	3.2997

X	$\log_{10}(X - 1990)$
1991	0
1992	.301
1993	.477
1994	.602

Transforming Skewed Distributions

- The example below shows how a $\log_{10}$ transformation can fix a positive skew
- The density estimate for average income for occupations from the Canadian Prestige data is shown on top; the bottom shows the density estimate of the transformed income

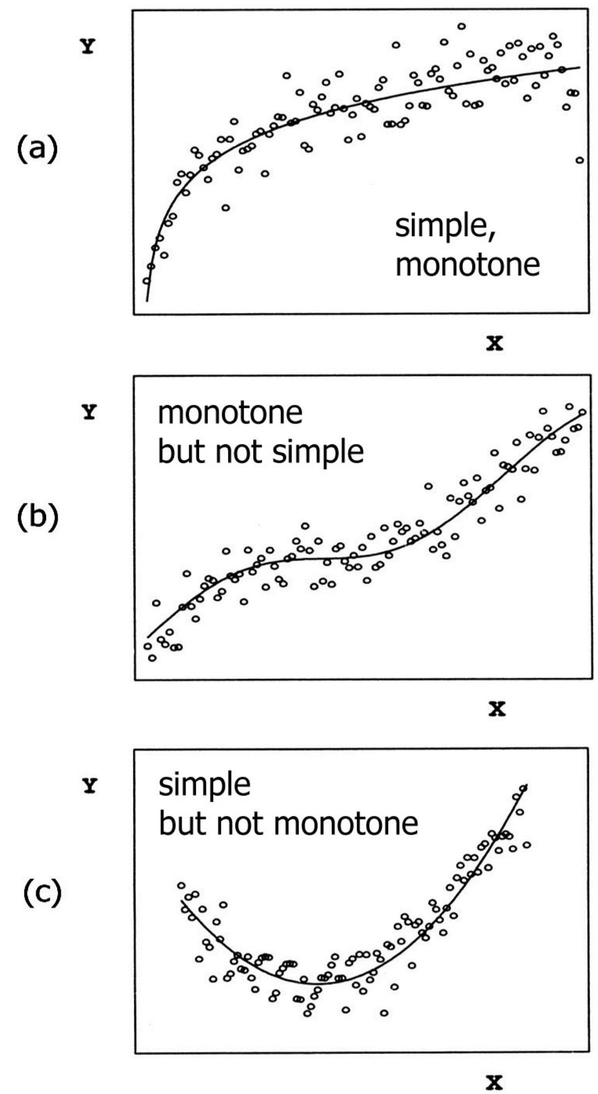
```
>log.inc<-log(income, 10)
>sm.density(income, model="normal")
>sm.density(log.inc, model="normal")
```



Transforming Nonlinearity When is it possible?

- An important use of transformations is to 'straighten' the relationship between two variables
- This is possible only when the nonlinear relationship is *simple* and *monotone*
  - Simple implies that the curvature does not change—there is one curve
  - Monotone implies that the curve is *always positive* or *always negative*
- (a) can be transformed, (b) and (c) can not

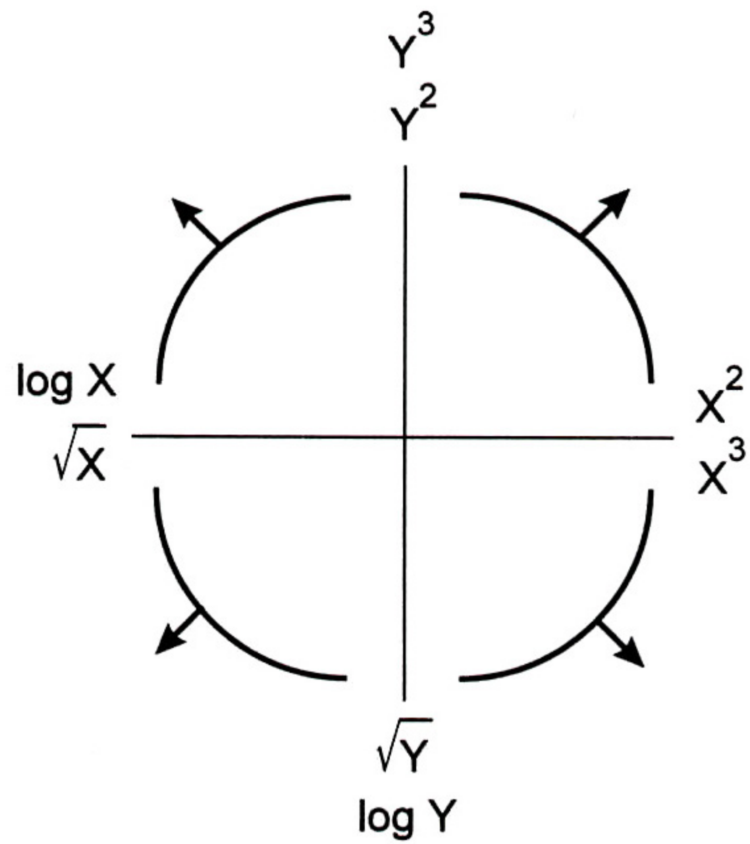
Figure 4.5 from Fox (1997)



The 'Bulging Rule' for transformations

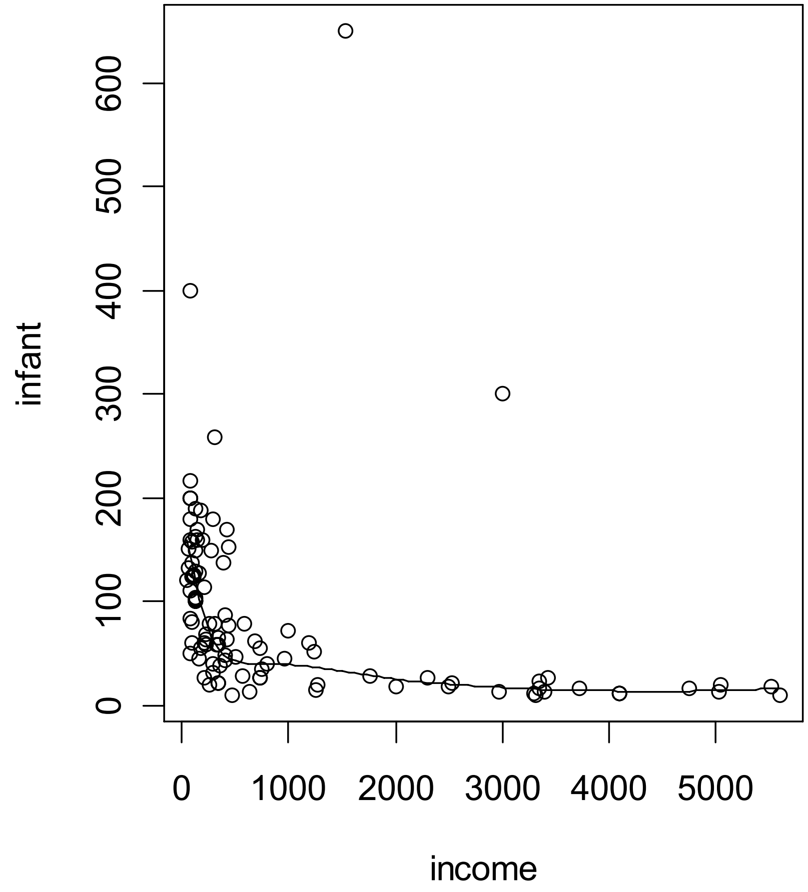
- Tukey and Mosteller's rule provides a starting point for possible transformations to correct nonlinearity
- Normally we should try to transform explanatory variables rather than the response variable Y since a transformation of Y will affect the relationship of Y with all Xs not just the one with the nonlinear relationship
- If, however, the response variable is highly skewed, it makes sense to transform it instead

Figure 4.6 from Fox (1997)



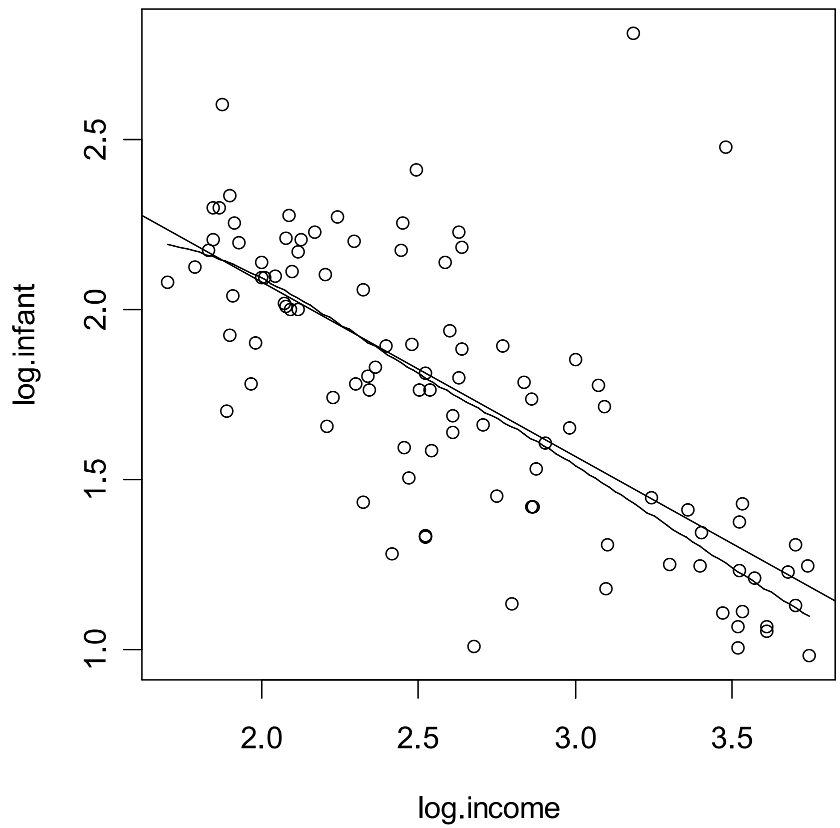
Transforming relationships Income and Infant mortality (1)

- Leinhardt's data from the `car` library
- Robust local regression in the plot shows serious nonlinearity
- The *bulging rule* suggests that both Y and X can be transformed down the ladder of powers
- I tried taking the log of income only, but significant nonlinearity still remained
- In the end, I took the $\log_{10}$ of both income and infant mortality



Income and Infant mortality (2)

Linear model after transformations				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1034	0.1375	22.57	0.0000
log.income	−0.5118	0.0512	−9.99	0.0000



- A linear model fits well here
- Since both variables are transformed by the $\log_{10}$ the coefficients are easy to interpret:
  - An increase in income by 1% is associated, on average, with a .51% decrease in infant mortality

Transforming Proportions

- Power transformations will not work for proportions (including percentages and rates) if the data values approach the boundaries of 0 and 1
- Instead, we can use the *logit* or *probit* transformations for skewed proportion distributions. If their scales are equated, these two are practically indistinguishable:

$$\text{logit} \simeq (\pi/\sqrt{3}) \times \text{probit}$$

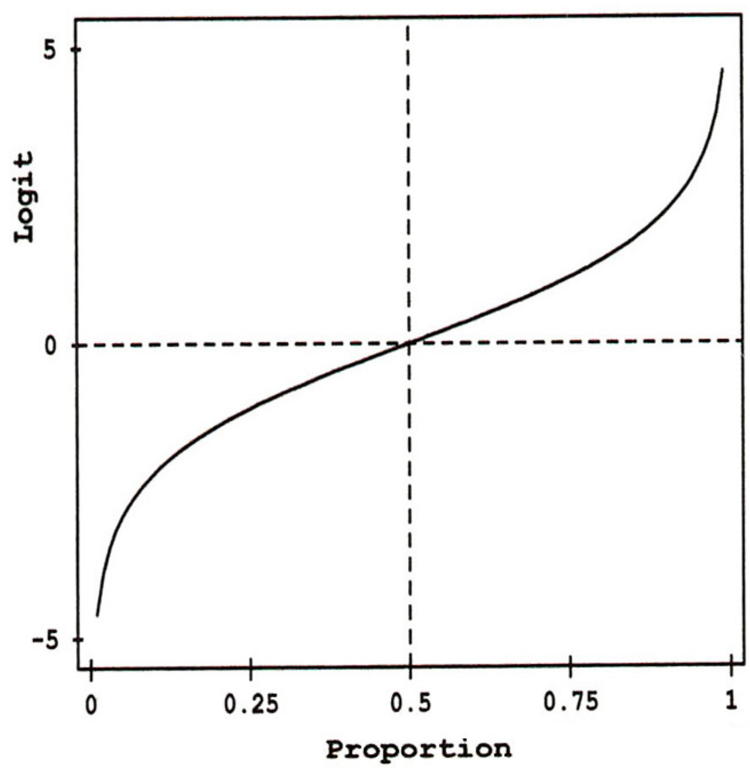
- The logit transformation: (a) removes the boundaries of the scale, (b) spreads out the tails of the distribution and (c) makes the distribution symmetric about 0. It takes the following form:

$$P \rightarrow \text{logit}(P) = \log_e \frac{P}{1-P}$$

Logit Transformation of a Proportion

- Notice that the transformation is nearly linear for proportions between .20 and .80
- Values close to 0 and 1 are spread out at an increasing rate, however
- Finally, the transformed variable is now centered at 0 rather than .5

Figure 4.15 from Fox (1997)




```
>library(car)
>logit(x) #logit transformation for a proportion
```

Next Topics:

- The Basics of Least Squares Regression
  - Least-squares fit
  - Properties of the least-squares estimator
  - Statistical inference
  - Regression in matrix form
- The Vector Representation of the Regression Model