

# Transformations

## READING:

## 1.4

## EXERCISES:

# CH.1: 27, 29ABC, 31, 32ABC

## ASSIGNED:

## HW 4

PRODUCER:

DR. MARIO



*IMG CREDIT: ALEX RIEGERT-WATERS*

# Reasons for Data Transformations

- Can Be Done on Response  $Y$  or Predictor  $X$
- Reasons for Data Transformations
  - Address **Non-Linear** Patterns
  - Stabilize Variance (Homoscedasticity)
  - Make Residuals More “Normally Distributed”
  - Minimize Effect of Outliers
- Focus is on **Non-Linear** Transformations

# Most Common Types

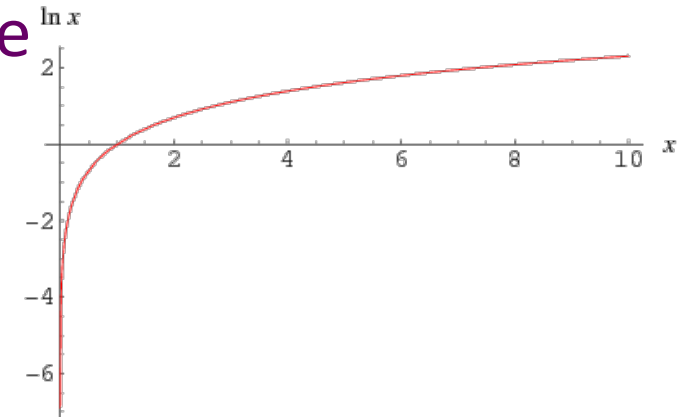
- Logarithmic:  $Y' = \log(Y)$
- Power Function  $Y' = Y^k$  where  $k \in \mathbb{R}$ 
  - Square Root (Special Case when  $k=2$ )
  - Reciprocal (Special Case when  $k=-1$ )
- Exponentiation:  $Y' = k^Y$ 
  - Exponential Growth  $Y' = e^Y$  where  $e = 2.71828$  (*Constant*)

# Choosing Transformations

- **Monotonicity:** Always increasing or decreasing
- **Easily Solvable for Y:** Need to Get Predictions For Y
- **Stronger Linear Pattern:** Look at Scatterplots
- **Better Residuals:** Look at Plots of Residuals
- **Box-Cox Transformation:** Helpful Method Not in Text

# Log Transformation

- Log Scales Seen in Science (Loudness, Earthquake, Acidity)
- Helpful With Converting Multiplicative to Additive
  - Rule:  $\log(ab) = \log(a) + \log(b)$
- Helpful With Converting Scales
  - Income is Always Positive (Linear Regression Appropriate?)
- Default in Statistics: Natural Log (Base  $e$ )



# Log Transformation

- Signs a Log Transformation Might Be Useful (For  $Y$  or  $X$ )
  - Variables are Highly Right Skewed
  - Values Range Over 2 or More Orders of Magnitude (Need Reference Value: Common is  $10^k$ )
  - Scatterplot Shows Logarithmic Growth
  - Plot of Residuals Show “Fanning Out” or “Megaphone”
  - Multiplicative or Proportional Relationships are Reasonable
- Easily Solvable:  $Y' = \log(Y)$  and  $Y' = \beta_0 + \beta_1 X + \epsilon$   
$$\hat{Y} = \exp(\hat{\beta}_0 + \hat{\beta}_1 X) = e^{(\hat{\beta}_0 + \hat{\beta}_1 X)} = e^{\hat{Y}'}$$

## *Make Reasonable Decisions*

