4. Transformations

Overview

- Linear least squares regression makes strong assumptions about the data:
  - Linear relation
  - Equal variance
  - Normal distribution
- Transforming the data can help satisfy these assumptions. It can also assist in examining the data.
- Disadvantage of transformations: interpretation becomes more difficult

Family of powers and roots

- Useful family of transformations: $X \mapsto X^p$
  - $p = 2$: $X \mapsto X^2$
  - $p = -1$: $X \mapsto 1/X$
  - $p = 1/2$: $X \mapsto \sqrt{X}$
- Little more complex, but easier to compare: $X \mapsto X^{(p)} = \frac{X^p - 1}{p}$.
- See picture.

Family of powers and roots

- Dividing by p is necessary to preserve the direction of X.
- All transformations match in value and slope at $X = 1$.
- We use the convention $X^{(0)} = \log X$ (because $\lim_{p\to 0} \frac{X^p - 1}{p} = \log X$).
- Ascending the ladder ($p > 1$) spreads out large values and compresses small values.
- Descending the ladder ($p < 1$) compresses large values and spreads out small values.

Method for finding transformation

- Method:
  - Use $X^{(p)}$ to find the right value of p.
  - Once you've found the right p, it is often easier to use X^p instead of $X^{(p)}$.
- Also, it is often easier to use ${}^{10}\log(X)$ or ${}^{2}\log(X)$ instead of the natural logarithm.

Using a 'start'

- If there are negative values, the transformation doesn't preserve direction $\rightarrow$ use a positive start.
- If the ratio of the largest to the smallest observation is close to 1 (≤ 5), then the transformation is nearly linear and therefore ineffective $\rightarrow$ use a negative start.
- We usually select values in the range $-2 \leq p \leq 3$, and simple fractions such as $1/2$ and $1/3$.
- Always keep interpretability in mind. If $p = .1$ seems best for the data, it is often better to use the log transformation ($p = 0$), because this is easier to interpret.

6 / 10

Transforming skewness

- Problems with skewed distribution
  - Data difficult to examine because most observations are in a small part of the range of the data. Outlying values in the direction opposite to the skew may be invisible.
  - Least squares regression traces the conditional mean of Y given the X's. The mean is not a good summary of the center of a skewed distribution.
- Right skew (positive skew) $\rightarrow$ need to compress large values $\rightarrow$ descend the ladder of powers $\rightarrow$ $p < 1$.
- Left skew (negative skew) $\rightarrow$ need to compress small values $\rightarrow$ ascend the ladder of powers $\rightarrow$ $p > 1$.
- See R-code.

7 / 10

Transforming nonlinearity

- Why do we want things to be linear?
  - Linear relationships are simple, and there is nice statistical theory for these models.
  - If there are several independent variables, nonparametric regression may be infeasible
- *Simple monotone* nonlinearity (direction of curvature does not change) can often be corrected using a transformation in the family of powers and roots
- Example: quadratic function - two possible transformations
- Mosteller and Tukey's Bulging rule
- Consider how transformation affects symmetry. If the dependent variable already was symmetric, then try to leave this one untouched. And again, keep in mind interpretability.
- See R-code.

Transforming nonconstant spread

- Differences in spread are often related to differences in level. Often: higher level $\rightarrow$ higher spread
- When spread is positively related to level, we need to compress large values $\rightarrow$ transformation down the ladder of powers and roots $\rightarrow p < 1$.
- When spread is negatively related to level (rare), we need to spread out large values $\rightarrow$ transformation up the ladder of powers and roots $\rightarrow p > 1$.
- See R-code.

Summary of transformations

- Advantage: transformations can help satisfy the assumptions of linearity, constant variance and normality.
- Disadvantage: interpretation is more difficult.
- The family of powers and roots (X^p or $(X^p - 1)/p$):
  - Ascending the ladder of powers ($p > 1$) spreads out large values and compresses small values.
  - Descending the ladder of powers ($p < 1$) does the opposite.