Transformation

- allows for the correction of non-normality caused by skewness, kurtosis, or other problems (lack of linearity)
- Shouldn’t be done if values represent meaningful scale
- Square root – moderate violations, LOG – severe, and inverse for severe violation
Transformation

- For positively skewed data square root and log keep data in the original order but bring in the spread, while inverse flips the order of the data.

- For negatively skewed data the reverse is true; without adjustment square root and log reverse order and inverse keeps the same order.
Original Data
Square Root Transform

![Bar Chart]

- Values: 1, 2, 3, 4, 5, 6
- Y-axis: 0, 2, 4, 6, 8, 10, 12
- Square root of values:
  - 1: 1
  - 2: 1.414
  - 3: 1.732
  - 4: 2
  - 5: 2.236
  - 6: 2.449

Note: The bar chart shows the square root transformation of the values 1 to 6.
Log Transform
Inverse Transform
Dealing with Missing Data

- The default in many programs (e.g. SPSS) is to do a complete cases analysis (listwise deletion)
  - simple and easy
  - but many concerns (e.g. percent of missing, pattern of missing) because doing complete cases analysis assumes missing at random
Missing Completely at Random

- MCAR – means that the patterns of missing on any one variable is not related to another variable.

- Example of non-MCAR: Measures of IQ and Income – subjects below a certain level of IQ (e.g. cutoff for “retardation”) may not have any income because they are under guardian care, so they leave the income variable blank.
# Complete Cases Analysis

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>$x_{11}$</td>
<td>$x_{12}$</td>
<td>$x_{13}$</td>
</tr>
<tr>
<td>$S_2$</td>
<td>$x_{21}$</td>
<td>$-$</td>
<td>$x_{23}$</td>
</tr>
<tr>
<td>$S_3$</td>
<td>$-$</td>
<td>$x_{32}$</td>
<td>$x_{33}$</td>
</tr>
</tbody>
</table>

Only this case is used.
Create a correlation matrix using complete cases for each pair of variables.

- For each correlation estimate you are using the most data possible.
- But each estimate is based on a different number of subjects.
- “Delete cases pairwise” in SPSS.
Dealing with Missing Data

- **Imputation (replacing missing data)**
  - Variable Mean insert – doesn’t affect the mean estimation but restricts the variance
  - Group mean insert – if you have grouped data then replace missing values with the mean of the group the subject belonged to.
  - Regression – predicting a subject’s missing value on one variable by scores on other variables. Could be used iteratively. Iterative means the process is repeated until the estimated value stabilizes.
Dealing with Missing Data

- **Imputation** (replacing missing data)
  - Estimation maximization (EM) algorithm – this is a maximum likelihood iterative estimation method.
  - Multiple Imputation – use multiple methods from above (and others in the book) and compute average estimate.
    - This is nice because it also gives you a standard error estimate for the estimation.