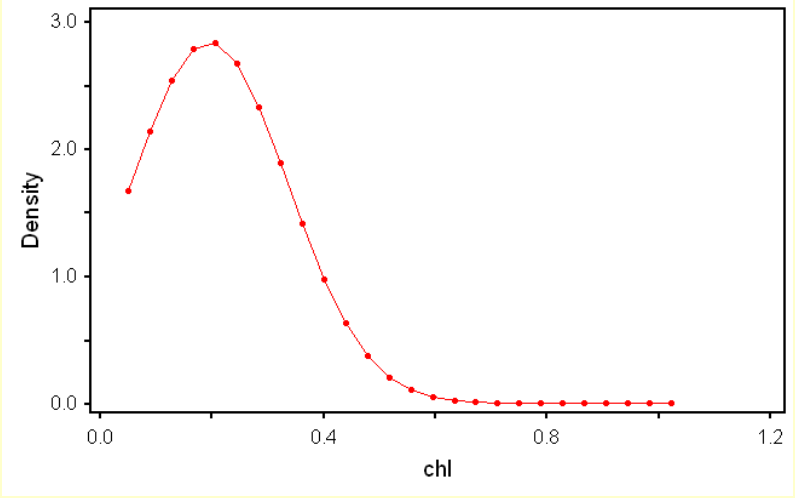
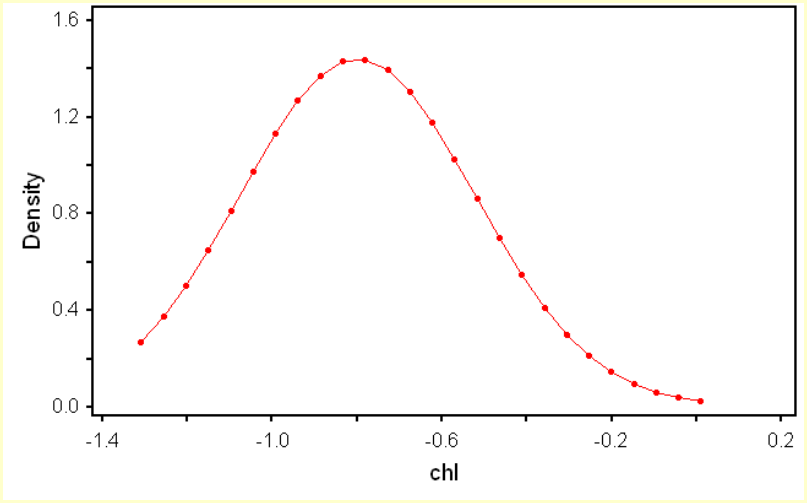
Data Distributions and Transformations





Chl Concentration → Log (Chl Conc.)

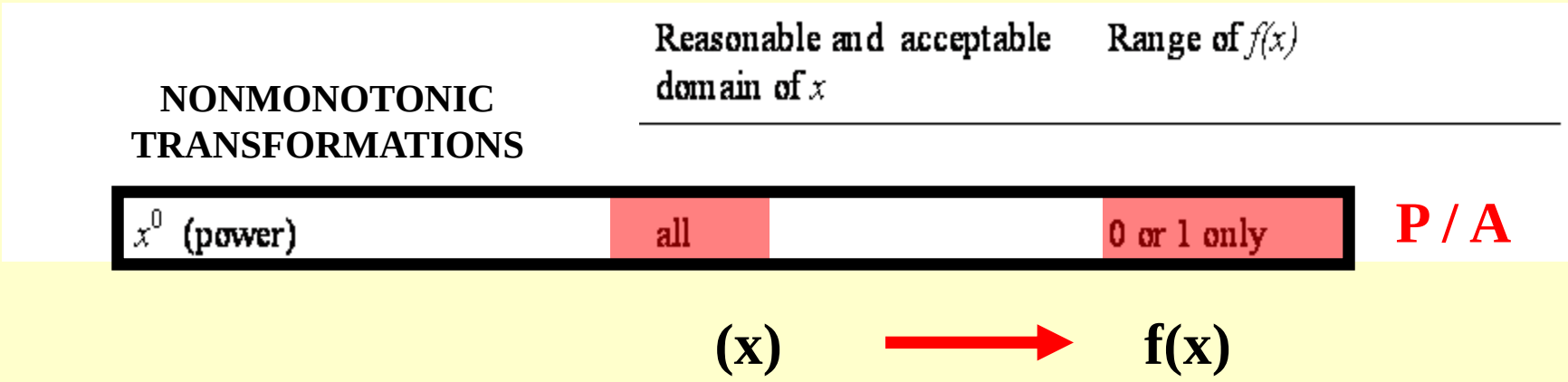
Approach

- We can transform the data to achieve normality
- Need to implement monotonic transformations:

  Actual Values Change

  Ranks Do Not Change

Presence / Absence Transformation



Note: 0 power transformation is NOT monotonic

It recodes data as Presence / Absence (0 / 1)

Log Transformation

Logarithmic transformation fx = ln(x) OR log(x)

TRANSFORMATION	Reasonable and acceptable domain of x	Range of $f(x)$
$\log(x)$	positive	all

(x) → f(x)

- This transformation is useful when:
  - High degree of variation within samples (e.g., Chl Conc.)
  - Large outliers (tails) and lots of zeros

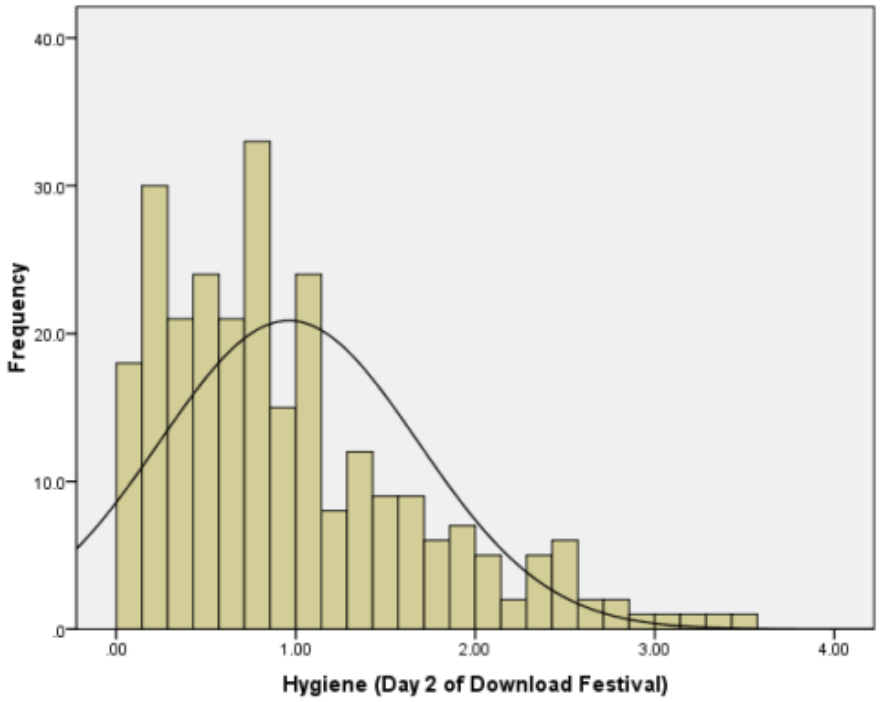
- Note: to log-transform data containing zeros, a small number should be added to all data points.

  - With count data, add one, so that: fx = log(0+1) =0
  - With density data, add constant smaller than smallest possible sample, so that: fx = log(0+0.001) = -3

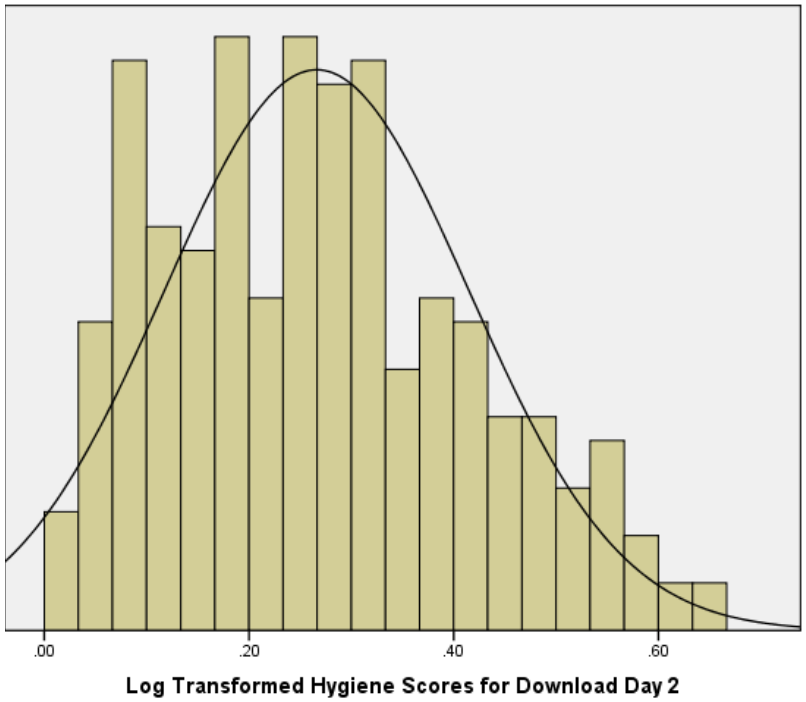
Log Transformation

Log Transformation ($\log(X_i)$): Reduces positive skew

Before



After



Square Root Transformations

MONOTONIC TRANSFORMATIONS	Reasonable and acceptable domain of x	Range of $f(x)$
$x^{1/2}$ (power)	nonnegative	nonnegative

Power exponents:

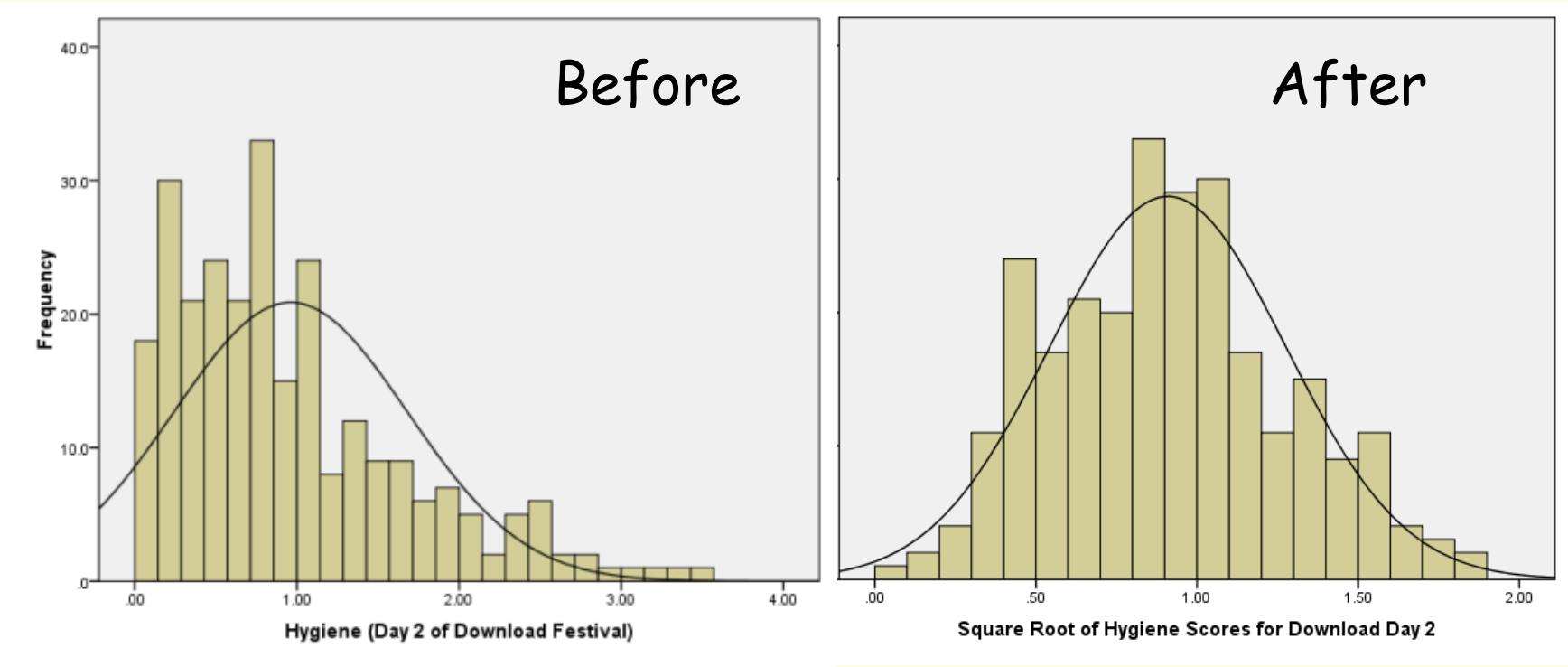
$\frac{1}{2}$ power (square root)

Square root transform deals with positive skew, by bringing in large tails.
Special treatment of zeros not necessary.

Square Root Transformation

Square Root Transformation ($\sqrt{X_i}$):

Reduces positive skew. Useful for stabilizing variance



Data Transformations – For Proportions

Arcsine / Arcsine-squareroot transformation

TRANSFORMATION	Reasonable and acceptable domain of x	Range of $f(x)$
$\arcsin(x)$	$0 \leq x \leq 1$	0 to 1 inclusive
$\arcsin(x^{1/2})$	$0 \leq x \leq 1$	0 to 1 inclusive

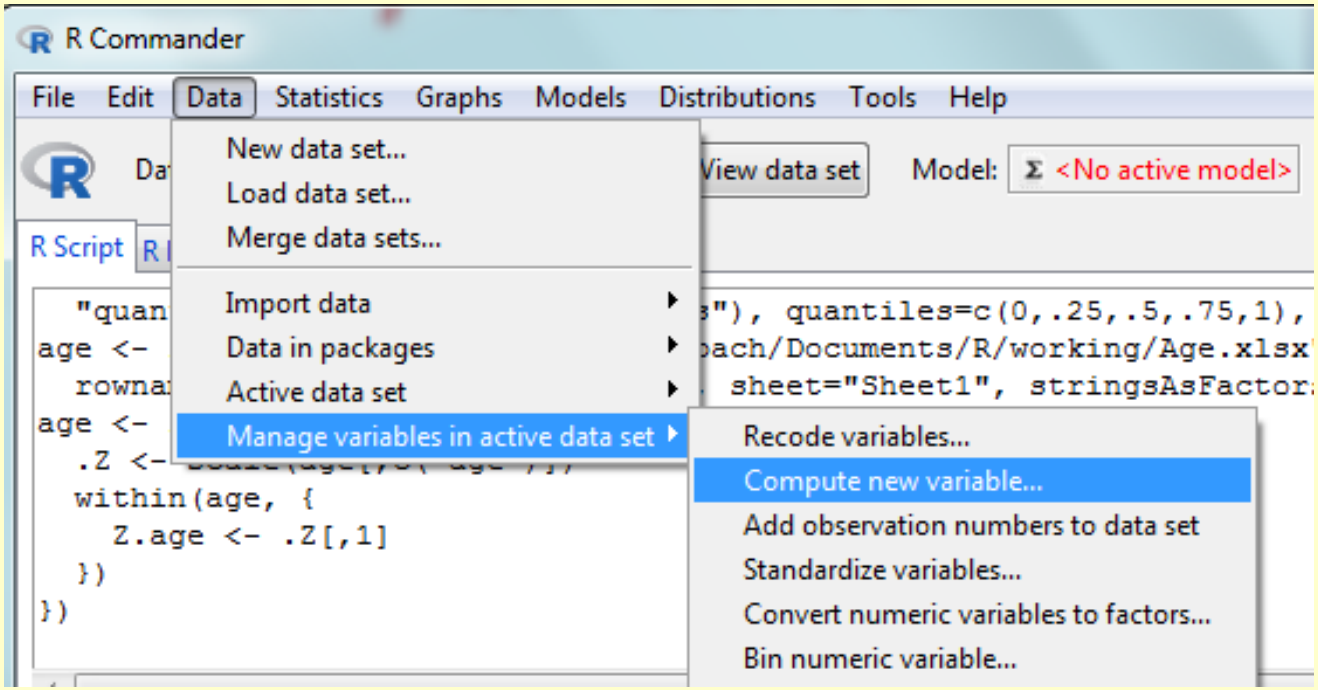
- This transformation is useful when dealing with proportional data (e.g., Percent Cover)

- Note: data must range between 0 and 1, inclusive.

The constant 2 / pi scales the result of arcsin(x) [in radians] to range from 0 to 1, assuming that $0 \leq x \leq 1$.

So, the Data are not Normal… Now What ?

Transform the Data in Rcmdr to Fix the Problem

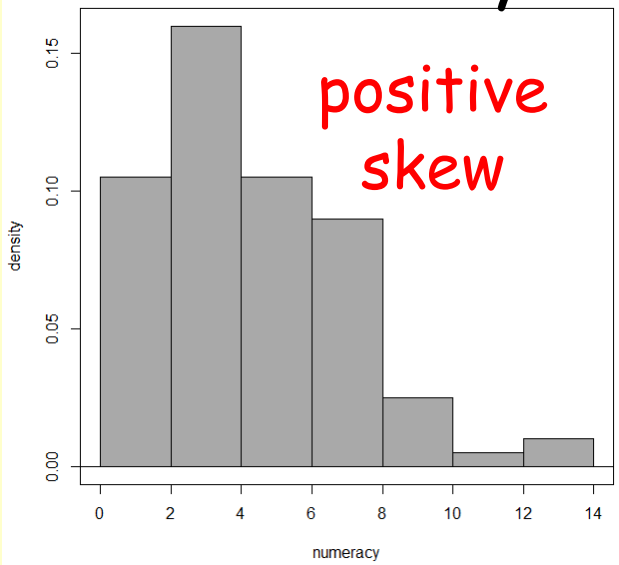


Data

Manage variables in active data set

Compute New Variable

Numeracy:



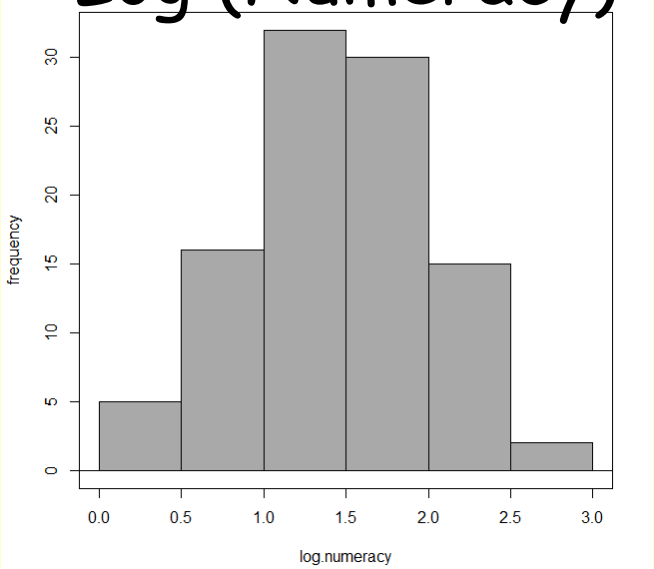
positive skew

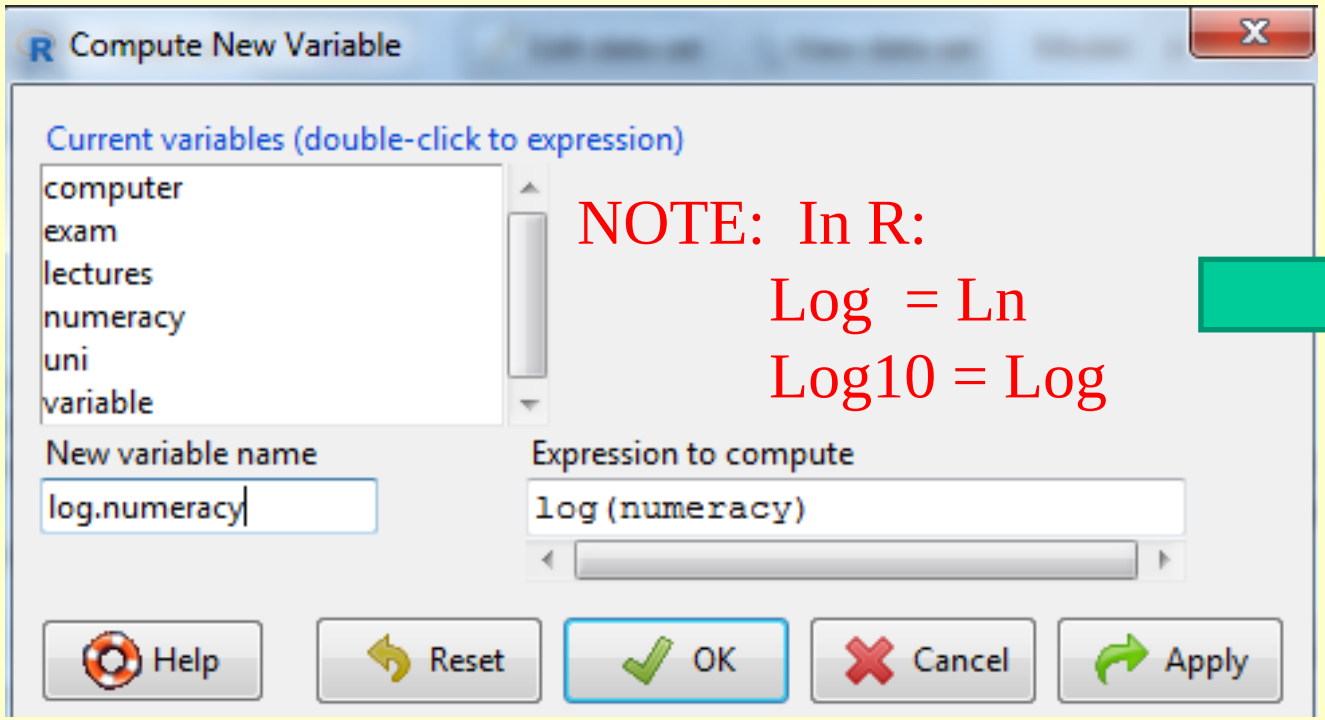
> Summary

Min: 1

Max = 14

Log (Numeracy):





NOTE: In R:

Log = Ln

Log10 = Log

log.numeracy
1.9459101
0.0000000
1.7917595
1.7917595
0.6931472
2.0794415
1.0986123
1.9459101
1.3862944
1.7917595
1.0986123
1.0986123
0.0000000

Hints for Computing New Variables

Log = is the natural ln

Log10 = is the log base 10

asin = arcsine

sqrt = square root

Summary

Rules for Data Transformations

Most Important Rule: Do not Reverse the Order of the Values (larger remains larger… smaller remains smaller)

Monotonic: change values but retain ranks

Non-monotonic: change values and ranks

(For example: Add random number, Multiply by (-1))

Summary
Take-home Lessons

- Parametric tests are more powerful, but are based on assumption of normally distributed data
- Determine normality criteria and undertake data transformations, if needed
- If you are unsure, data transformations can always be attempted to compare the same test results, using transformed and un-transformed data
- Test normality before / after data transformations
- If transformations do not work…

  use non-parametric tests