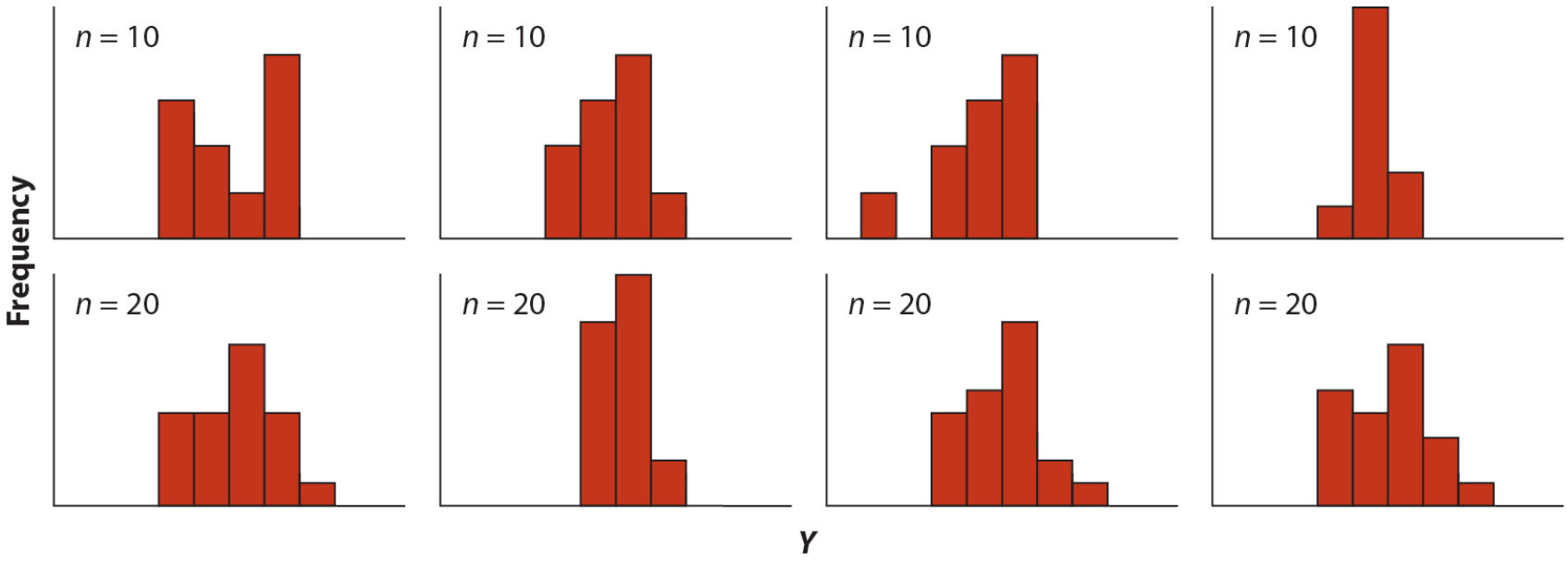
Data transformations and Non-Parameteric Methods

Paul M. Magwene

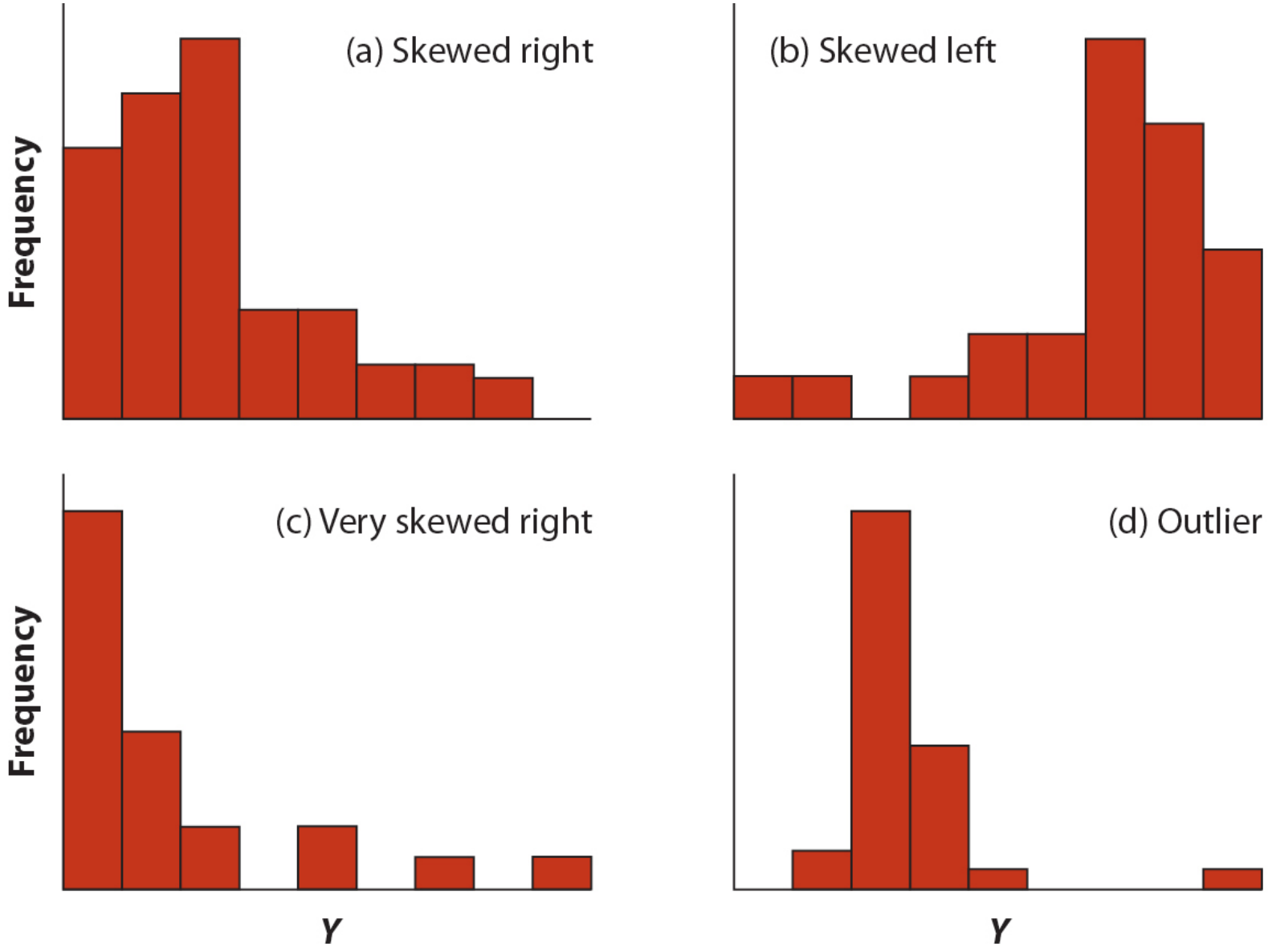
07 December 2016

What do you do if your data is not normally distributed?

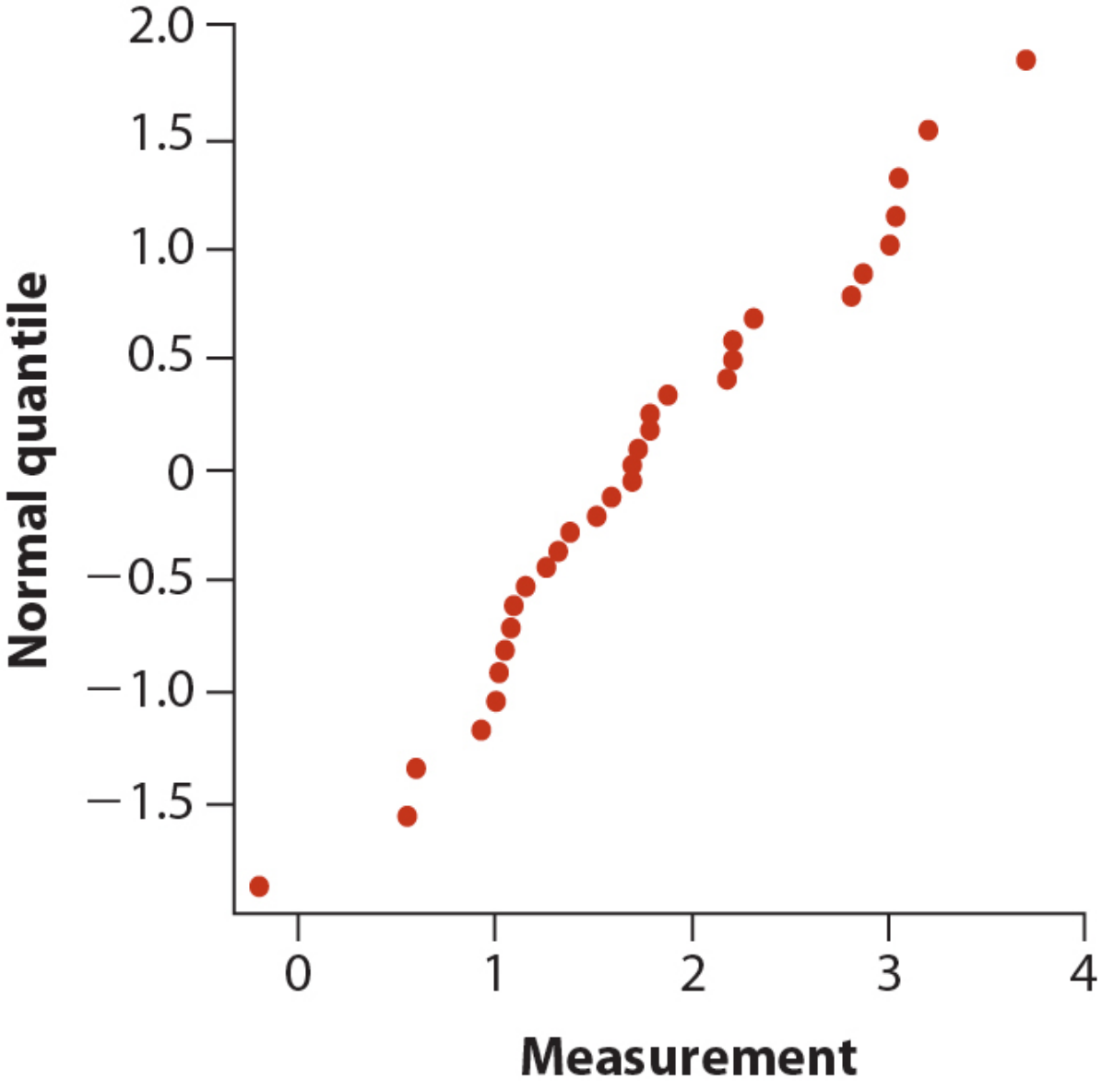
Is my data normally distributed?



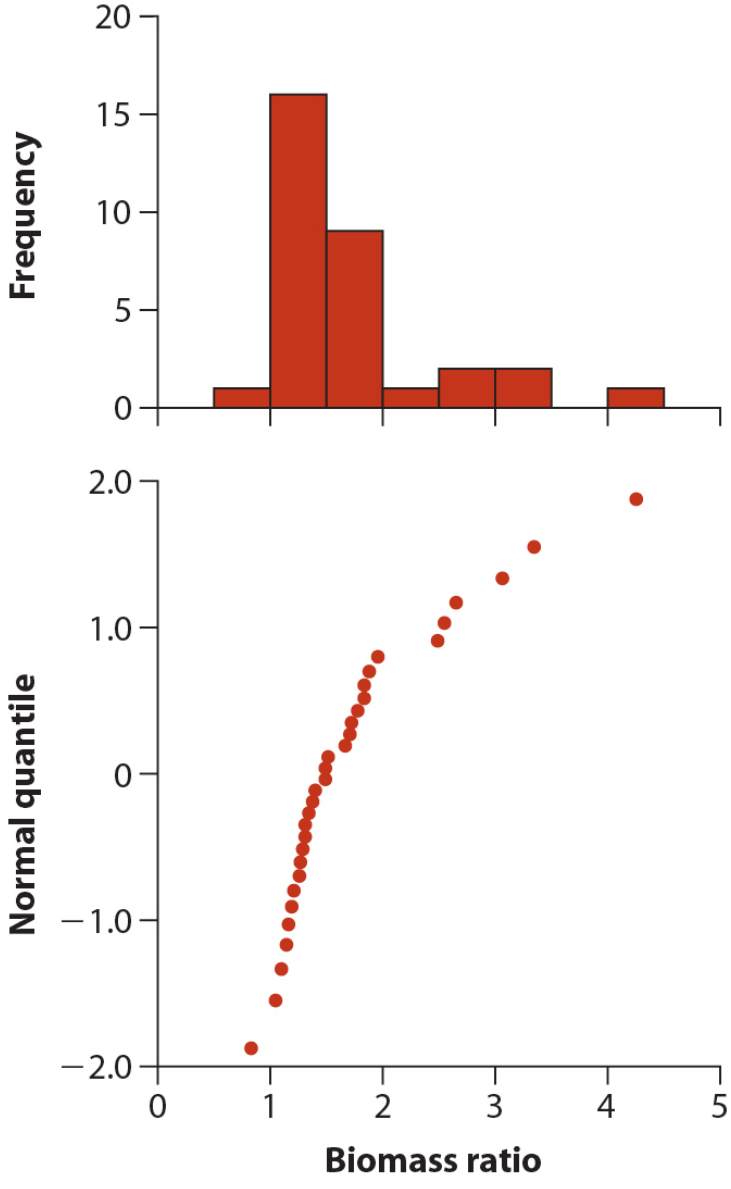
Some non-normal distributions



Visual tools: normal quantile plot



Example data set: Comparing biomass between protected and unprotected marine sites



A formal test for normality: Shapiro-Wilk Test

Data transformations

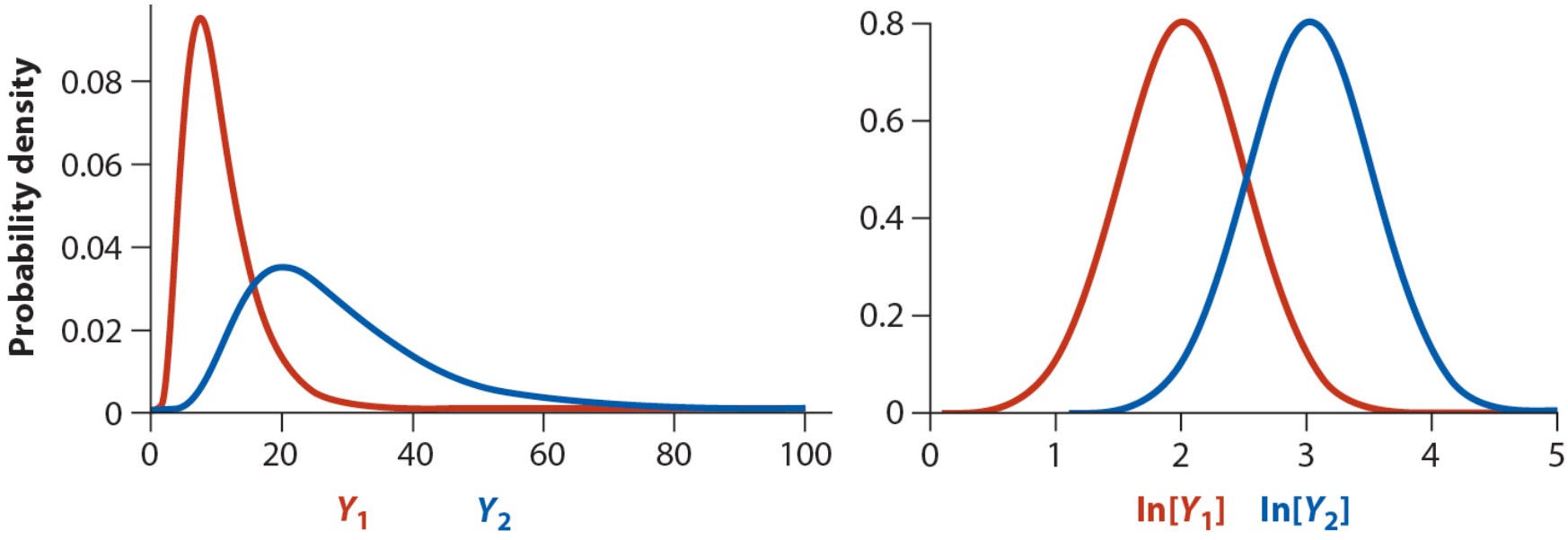
Log transformation

$$X' = \ln[X]$$

Tends to work well when:

- The data are all positive
- The frequency distribution is right skewed
- The data span several orders of magnitude
- The measurements are ratios or products of variables
- e.g. morphological measures such as body mass, length

Log transformation, cont.



Log transformation of biomass ratio data set

Cautions

- $\bar{X}' \neq \ln[X]$
- Often will do analyses in log transformed data, and then back transform to original scale to report *geometric mean* and CIs to facilitate interpretation

Arcsine transformation

$$X' = arcsin[\sqrt{X}]$$

Used when data are proportions

Square-root transformation

$$X' = \sqrt{X + 1/2}$$

Used when the data are counts, such as number of eggs laid, number of bacterial colonies, number of mates acquired, etc.

Other transformations

Square transformation, $X' = X^2$

- left skewed data

Natural exponential function, $X' = e^X$

- alternative for left skewed data

Reciprocal transformation, $X' = \frac{1}{X}$

- right skewed, all data points have the same sign

Non-parametric tests

Sign test (alternative to one-sample t-test)

Non-parameteric alternative to one-sample t-test

Mann-Whitney U-test (alternative to two-sample t-test)

Non-parameteric alternative to two-sample t-test

Kruskal-Wallis test (alternative to ANOVA)

Spearman's rank correlation

Assumptions of non-parametric tests

Power of non-parametric tests