Proportions, percentages, ratios

(1) **Proportion**: fraction having a denominator, $N$, that is *fixed* either by nature or by the experimenter:

\[
p = \frac{X}{N} \quad 0 \leq X \leq N
\]

\[
p = \frac{0.01}{0.01} \quad 0 \leq p \leq 1
\]

- $X$ is often a count, but may be a continuous variable.
- $N$ is fixed:
  - by the experimenter: e.g., number of heads per 10 coin flips; number of plants per 10 m$^2$ quadrat;
  - by nature: number of days of rainfall per year.
- Proportions are very well behaved statistics.
**Proportions**

*Example:* number of heads per 100 flips; experiment is repeated 1000 times.
Sampling distributions of proportions

- Only problem:
  - Because the proportion has fixed bounds, sampling variance is a function of the mean:
Sampling distributions of proportions

- **Transformations** that correct this problem:
  - Good: the “angular” or “arcsine” transformation:
    \[ p' = \text{arcsin} \sqrt{p} \]
  - Better: the Freeman-Tukey correction (1950):
    \[ p' = 2\sqrt{N} \left[ \text{arcsin} \left( \frac{X + \frac{3}{8}}{N + \frac{3}{4}} \right)^{\frac{1}{2}} - \text{arcsin} \sqrt{p} \right] \]

- Note: these transformations are defined for **proportions**, not **ratios**.
Sampling distributions of proportions

1. Frequency of proportions

2. Frequency of arcsine-transformed proportions

3. Frequency of proportions

4. Frequency of arcsine-transformed proportions
Proportions, percentages, ratios

(2) Percentage: $100p$

- All well-behaved: have same distributions as proportions.
- Commonly used to report results, but care must be taken because of shifting reference points.

- Your boss says to you: “Our company is having temporary financial difficulties in meeting a loan payment, so I must temporarily give you a 50% cut in salary. But don’t worry, because next month I’m going to give you a 90% pay increase.” So in the long run you’ll be sittin’ pretty, right?

- A news report: “Last year there was a 10% decrease in stockpiles of nuclear weapons. This past year there was another 10% decrease, for a total of 20% in two years.” Right?
Proportions, percentages, ratios

(3) **Ratios:** have variability both in the numerator and in the denominator:

\[ R = \frac{X}{Y} \]

E.g.: \[ RHL = \frac{\text{Head length}}{\text{Body length}} \]

- \( X \) and \( Y \) are usually positive.
  - \( Y \) must be non-zero.
  - Both are subject to variation.
- Ratios are widely used as *derived variables*.
  - Purpose is to *remove or downplay the effects of size or magnitude* by dividing by a standard variable.
  - Conceptually *intuitive*.
  - Called an *index* when a series of measures is divided by a common standard variable (usually the largest).
- *Ex:* body proportions, sex ratios, physiological ratios, isotope ratios, edaphic indices, diversity indices, prey-preference indices.
Important problems with ratios

(a) **Loss of information:**

- If the ratio increases, it might be because the numerator increases, the denominator decreases, or some combination of both.

\[
R_0 = \frac{2}{2} = 1, \quad R_1 = 2
\]

\[
R_1 = \frac{4}{2}?
\]

\[
R_1 = \frac{2}{1}?
\]

\[
R_1 = \frac{3}{1.5}?
\]
Important problems with ratios

(b) **Decreased precision** due to *compounded errors*:

- E.g., $X = 1.2$, $Y = 1.8$, both measured to $\pm 0.05$.
- Using the min-max procedure:
  - $X$: range 1.15 to 1.25, max 4.3% error.
  - $Y$: range 1.75 to 1.85, max 2.9% error.

  - $X + Y$: range 2.90 to 3.10, max 3.3% error.
  - $\frac{X}{Y}$: range $\frac{1.15}{1.85}$ to $\frac{1.25}{1.75}$, max 7.0% error.

- Ratios compound sampling errors.
Important problems with ratios

(c) Ratios tend to have curious sampling distributions:

- Highly asymmetric: right-skewed and peaked.
Important problems with ratios

- “Ideal” case in which $X$ and $Y$ are both normally distributed with identical variances:
  - Ratio has a Cauchy distribution.
  - No parameters: no mean, variance, etc.
- Complicates the estimation of standard errors of indices, which are often very complex in structure.
- Such distributions can mask real differences among groups.
Important problems with ratios
Important problems with ratios

(d) **Division by a common variables** (e.g., indices) can introduce *spurious correlations* among otherwise independent variables (Karl Pearson, 1897).

- Division of two independent variables by a third variable can introduce a common source of information.

- The greater the magnitude (and variance) of the denominator, the greater the spurious correlation.

- Particular problem with biological indices, which often use the largest variables (e.g., body length) as the denominator for all ratios.
Important problems with ratios

Correlation = 0.08

Correlation = -0.29

Correlation = -0.05

Correlation = 0.77
Sampling distributions of correlations of ratios

- N = 10
- N = 25
- N = 50
- N = 100

Correlation of X/Z and Y/Z
Important problems with ratios

(e) Except in very special circumstances, ratios do not adjust for magnitude:

- Ratios tend not to be independent of the numerator and denominator.
- For example, assume the simplest case: a linear relationship between $X$ and $Y$:

  $$ Y = mX + b $$

  $$ \frac{Y}{X} = \frac{mX + b}{X} = \frac{mX}{X} + \frac{b}{X} $$

  $$ \frac{Y}{X} = m + \frac{b}{X} $$

- Ratios are independent of size only if $b = 0$.
- Even in that case, ratios change the correlation structure of the data because they change the distributions.
Important problems with ratios
Indices revisited

• Hundreds of indices are used in the biological (and other) sciences.
• May involve simple ratios, but are often more complex:
  – Ratios of ratios.
  – E.g., diversity indices.
• Some well-grounded in distribution theory, but most are *ad hoc*.
• Ratios and indices are often used for summarizing results, but should *never* be used as variables in a statistical analysis.
• What to do instead of using ratios?

(1) If working with a previously collected data set and have only ratios:
  • Apply logarithmic transformation (to any base).
  • Transforms multiplicative relationship between numerator and denominator to an additive relationship:
    \[
    \log \left( \frac{X}{Y} \right) = \log X - \log Y
    \]
  • Will correct distributional problems, but not problems of loss of information nor of spurious correlations.
(2) For *univariate* analyses (e.g., ANOVA) of ratios:
- Use “nonparametric” equivalent analyses.
- E.g., Mann-Whitney or Kruskal-Wallace tests.

(3) For multivariate analyses of ratios:
- Randomization methods must be used.
- Resulting hypothesis tests will usually be less powerful than comparable tests with original data.

(4) In general:
- No reason to use ratios in morphometric analyses.
- Always use original measured variables.