Statistical Science

Chapter 2.1 Five part definition of a measured quantity

ReCap (Ch 1)
Quantities (Ch2)
2.1 Five part definition
2.2 Types of measurement scale
2.3 Data collection, recording, and error checking
2.4 Graphical and tabular display of data
Critique of graphs and tables (Lab 5)
2.5 Ratio Scale Units
Base
Standard Multiples
Commonly used units in biology
2.6 Dimensions

on chalk board

```
Not here last time?
     Syllabus
     Name on roster

Lab 1
 Bring Cards
 Location: cf syllabus
```

```
Break, if no quiz:
2 copies of 'Quantitative
Answers to Rhetorical
Questions'
```

```
Ch2.1  needed for
problem set 1
```

ReCap Chapter 1

The Role of Statistics in Science

Statistics have come to play a central role in the biological, psychological, and health sciences.

Model Based Statistics in Biology

Simplification required to deal with uncertainty and with biological complexity
$\triangle$ Verbal, graphical, and formal model (equations)

Models are used to make: useful calculations (species extinction rate)
decisions about experiments (yes/no)

Role of statistics: Development of models (exploratory analysis)
Formal evaluation of models (confirmatory analysis)

Quantitative reasoning about biological and environmental phenomena.

It is not a course in math. It is not a course in rote learning of statistical tests.
It is a course in how to think with measured quantities.
It will integrate models with statistics.

Wrap-up

Course will integrate models with statistics.
Models express ideas about the relation of quantities.
Measured quantities are defined in 5 parts.

Quantities (from Schneider 2009, Quantitative Ecology Chapter 3)
Scientists work with measured quantities, not with numbers or mathematical abstractions divorced from measurement (Riggs 1963).
Biologists work with quantities that have names and scaled values:

a density [N] of 5000 animals per hectare,
an increase rate r of 4% per year,
a mutation rate μ of 10^{-6} per generation.

Our interest is in biologically interpretable quantities, not the mathematical manipulation of symbols.

When told that $\frac{dx}{dt}$ means $\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$

The physicist Kelvin exclaimed
"Does nobody know that it represents a velocity ?" (Hart 1923)

One can open many journals and find equations with undefined symbols, symbols that represent measurements as numbers only, or measurements that lack units. Why bother with a definition of a measured quantity? There are several reasons.

One reason is communication. Numbers and symbols with no units result in irreproducible science. Without defined quantities we cannot compare one study with another.
A second is linking measurement to theory at appropriate scales. Measurements at appropriate scales. A theory that relates respiration rate to body size applies only within the range of the measured quantities.
A third reason is different rules. The rules for scaled quantities differ from those for numbers. The rules for scaled quantities differ from those for algebraic manipulation of symbols free of units and dimensions. One can take the logarithm of the number 4, one can take the logarithm of a parameter with a value of 4 (if it has no units), but one cannot take the logarithm of 4 mosquitoes. The symbols A and B can be added, 4 can be added to 8, but 4 mosquitoes cannot be added to 8 elephants. No-one would add 4 and 20 blackbirds to the number π = 3.14 the expression N + π makes it all too plausible unless quantities are defined and distinguished from numbers.

Quantities -- Definition
A well-defined biological quantity has 5 parts:
a name;
a procedural statement that prescribes the conditions for measurement, or calculations from measurements;
a set of scaled numbers generated by the procedural statement;
units on one of several types of measurement scale;
a symbol that stands for the set of scaled numbers.

Five part definition of a Quantity.
Here is an example.

PROCEDURAL STATEMENT	NAME	SYMBOL	VECTOR OF OUTCOMES	SCALE
(gravimetric mass, just after pupation, *etc*)	pupal mass	pM =	$\begin{bmatrix} 280 \\ 250 \\ 300 \end{bmatrix}$	milligrams

The procedural statement must have enough information that somebody else could take data that were comparable. For example, the statement of *pM* should state the stage of the pupa, species of insect, time of year.

Note: Some authors add a 6th part, the uncertainty, defined as half the resolution. In this example the uncertainty is half a milligram, assuming measurement to the nearest milligram.

Another example

PROCEDURAL STATEMENT	NAME	SYMBOL	VECTOR OF OUTCOMES	SCALE
number of heart beats per minute recorded with pressure sensor	heart rate	f_{heart} =	$\begin{bmatrix} 70 \\ 80 \\ 75 \end{bmatrix}$	min^{-1}

Heart rate is a frequency, hence the symbol *f* subscripted by heart. Like other measured quantities, this quantity is confined to a range. For healthy adults the range is 60 – 100 beats per minute, with an average around 75 for 20 year olds. Equivalently, heart beat frequency ranges from 1 to 1.67 Hz, where 1 Hertz = 1/sec, the standard unit for periodic phenomena. Compare this to other periodic measures in our lives in the1-2 Hz range.

	1 Hz	2 Hz
Locomotion	walking	running
Musical rhythm	Ballad	Rock and Roll
Metronome setting	Andante	Allegro
Period of a pendulum	0.25 m	1.1 m

Leg length is approximately 50 percent of total stature to achieve the biomechanical efficiency of the human striding bipedal gait. A leg swings like a pendulum for people with legs ranging from 0.7 m (1.4 m person) to 0.92 m (1.8 m person), at frequencies of 1.6 to 1.75 Hz respectively.

Quantities -- more about each part

Procedural Statement

These are typically found in the "methods" section of article in a scientific journal.
It must contain enough information so another person can use it to obtain comparable numbers.
This requires clear writing sufficient detail to be reproducible.

Derived quantities.
These are calculated from two or more measurements in standard units. For example density is the ratio mass to volume (kg m^{-3}). Specific gravity is the ratio of the density of a sample to that of pure water at a 15° C. Salinity (parts per thousand ppt) is calculated from hygrometer readings of specific gravity.

Derived quantities are measured by calibrating to standard units. For example, salinity is easily and accurately measured by conductivity, than applying a conversion factor to obtain salinity. The procedural statement usually assume these can be looked up if needed.
In cases of ambiguity (*e.g.* biodiversity) and equation is useful. For example the Shannon Wiener index for a sample with i =1,2 ....n species is

$H' = \Sigma p_i \ln p_i$ where p_i is the proportion of each species

Names

1. Read in words

Quantities should be read as names ('per capita birth rate') not as symbols ($\dot{N}$) because the name conveys more meaning. Symbols appear in prose for the sake of preciseness, or in mathematical expressions for the sake of clarity, but should still be read as names.

2. Associate the name with mental picture

Facility in reasoning with quantities comes in associating together the name, the symbol, a mental image of the biology. For example the quantity 'per capita fledging rate' is associated with a symbol $\dot{N}$ and with an image of a dynamic quantity, such as chicks jumping out of a nest each year. It is of note that mathematical notation uses the same part of the brain as recalling names and faces.

3. Associate name with a typical value

Measured quantities usually cluster around a central value. The quantity takes on meaning when the name, symbol, and image are further associated with a typical value obtained from simple calculation. An example is fldging rate $\dot{N} = \log_e$(5 chicks/2 parents)/year = 92%/year).

<u>Symbols</u> -- Choice

Skilful choice aids in quantitative work.
Attention to good choice is a neglected art in biology.
Here are some criteria for selecting mathematical notation.

Criteria in selecting mathematical notation. Modified from Cajori (1929). Table 7.2 in Schneider 1994, 2009, p131.

```
1.  Consistent with current usage.
2.  Reduces the burden on memory.
3.  Demonstrated utility in quantitative work.
4.  Brevity.
5.  Lends itself to computer applications.
```

Quantities – Criteria in selecting mathematical notation in science.

1. Use conventional symbols, consistent with current use.
   For example use g for gravitational acceleration.
   Use H' for Shannon-Weaver species diversity.
   Mathematical notation is a language.
   Hence the need to hold to conventional usage.

2. Mnemonic--easy to recall.
   For example, M for mass of an organism, f for frequency of heart beat.
   Coordination between symbols aids recall. Example: x y z for 3 directions in space.
   Diacritical marks aid recall.
   An example, common in physiology, is to place a dot over a symbol to represent the time rate of change.
   If E is energy (Joules) then $\dot{E}$ is energy exchange (Joules/sec)
   Diacritical marks are used in statistical notation.
   Place a bar over a symbol to designate the mean.
   For example use $\bar{A}$ rather than m_A to designate the mean surface area of lakes in a district. This establishes area as the quantity of interest, with the mean as a secondary value, derived by calculation.
   Another example: $\hat{\beta}$ is the estimate for the parameter β
   Diacritical marks bring out the relation between quantities.
   Icons instead of letter symbols. Suggested by a fisheries scientist, John Pope.
   For example rate of change in bird number: *d* 🕊 / d*t*
   (1/🕊) *d* 🕊 / d*t* "per capita rate of change in bird numbers"
   Do these aid recall ? What do you think ?

Quantities – Criteria in selecting mathematical notation in science.

3. Demonstrated utility. Example of changing use of dot notation.
   Newton's dot notation for rate of change in a quantity was initially useful.
   It was eventually dropped by mathematicians.
      It was cumbersome for 2^{nd} and 3^{rd} derivatives.
      It did not capture the mathematical concept of taking the limit.
   The dot notation was adopted by physiologists: $\dot{E}$ for time rate of change in energy E

4. Brevity. Simplify recurrent groups of symbols.
   For example, use r for the intrinsic rate of increase,
      rather than using $\mathrm{d}\ln N / \mathrm{d}t$

   Another example $$\vec{N} = \frac{1}{A}\frac{dN}{dt} \quad \textit{the horizontal flux of } N$$

5. Lend themselves to computer application.
      Related to this is Cajori's (1929) criterion of lending itself to typesetting.
      However, standard keyboards are required when writing formulae for computer calculations.

Quantities -- Skilful choice of symbols contributes to understanding quantitative ideas

Symbols -- Usage

Consistent usage is just as important as skilful choice.
   Meaning must remain consistent. Same procedural statement and units.
   It is all too easy for a name to stand for several different quantities,
      each with different procedural statements, or even different units.
   Symbols tied to names and units useful in distinguishing quantities
      that appear to be the same because they have the same name.
   Using names (with no symbols) leads to sloppy practice, because the
      different quantities can have same name.

Dictionary helps.
   If there are more than a few symbols, they should be collected together with
      definition in words, so that they can be looked up.

Concordance also helps.
   Write out full equation using parallel notation.
   Example. Notation in this course uses β notation for comparing slopes and means.
   Texts that do not emphasize a model based approach use different notation.

References

Cajori, F. 1929. *A History of Mathematical Notations*. Chicago: The Open Court Publishing Company.

Hart, I.B. 1923. *Makers of Science*. London: Macmillan.

Riggs, D.S. 1963. *The Mathematical Approach to Physiological Problems*. Baltimore: Williams and Wilkins. 1977 reprint by M.I.T. Press, Cambridge, Massachusetts.

Schneider, D.C. 1994, 2009. *Quantitative Ecology. Spatial and Temporal Scaling*. San Diego: Academic Press.