Model Based Statistics in Biology

Chapter 2.1 Five part definition of a 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

$\underline{\Delta}$ 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 phenomena.

Not a course in math. 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.
Quantities defined in 5 parts.

Quantities (from Schneider 2009, Quantitative Ecology Chapter 3)
Biologists, like all natural and social scientists, work with definable 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 $\dfrac{dx}{dt}$ means $\lim_{\Delta t \to 0} \dfrac{\Delta x}{\Delta t}$

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

One can open any biological journal and find equations with undefined symbols, symbols that lack biological interpretation, and 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 that defined quantities link theory to measurement. Defined quantities allow us to test theories about the relation of one quantity to another with appropriate measurements at appropriate scales.
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 was 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} =	$\begin{bmatrix} 45 \\ 50 \\ 46 \end{bmatrix}$	min^{-1}

Quantities -- more about each part

Procedural Statement

Typically found in the "methods" section of article in a scientific journal

Enough information so that another person can use it to obtain comparable numbers.

Written clearly.

Often include equations to show how one quantity was calculated from another.

Example a measure of the diversity of a sample with i =1,2 ....n species
$H' = \Sigma p_i \ln p_i$ where p_i is the proportion of each species
An equation is used to define this quantity.

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 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 birth rate' is associated with a symbol $\dot{N}$ and with an image of the quantity, such as chicks jumping out of a nest each year.

3. associate with a typical value

Name, symbol, and image are further associated with a typical value obtained from simple calculation ($\dot{N} = \log_e$(5 chicks/2 parents)/year = 92%/year).

Symbols -- Choice

Skilful choice aids in quantitative work.
Attention to good choice is a neglected art in biology.

Criteria for selecting mathematical notation.

Table 7.2 (Schneider 2009, p131)

Criteria in selecting mathematical notation. Modified from Cajori (1929).

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 -- more about each part

<u>Symbols</u> -- Choice

1. Conventional. For example use g for gravitational acceleration.
   Mathematical notation is a language.
   Hence need to hold to conventional usage.

2. Mnemonic--easy to recall.

Coordination between symbols aids recall. Example: $x\ y\ z$ for 3 directions in space.

Diacritical marks aid recall.

Place a bar over a symbol to designate the mean. $\bar{A}_{xy}$ rather than m_A to designate the mean value of the surface area of lakes in a district.

Another 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 bring out the relation between quantities, as in these two examples.

Icons aid recall (?). *d* 🕊 / d*t* due to John Pope. Explain

(1/🕊) *d* 🕊 / d*t* "per capita rate of change"

better ? no better than letters ? What do you think ?

3. Demonstrated utility. Example of changing use of dot notation.
   Newton's dot notation for rate of change in a quantity initially useful.
   Eventually dropped by mathematicians. Cumbersome for 2nd 3rd deriv.
   Did not emphasize mathematical concept of taking the limit.
   Then adopted by physiologists, for example, $\dot{E}$

4. Brevity. Simplify recurrent groups of symbols.
   For example, use r for the intrinsic rate of increase,
   rather than using d ln N / dt

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

5. Lends itself to computer application.
   Related to this, Cajori's (1929) criterion of lending itself to typesetting.
   Still use keyboards to write formulae to computer.

Skilful choice of symbols contributes to understanding of quantitative ideas.

Quantities -- more about each part.

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 model based approach use different notation.