# EEB 2208E: TOPIC 2

## INTERPRETING STATISTICS (WHEN THERE IS AN AGENDA)

## Background for this topic

Sutherland et al. 2013. Twenty tips for interpreting scientific claims. Nature 503:335-337. http://www.nature.com/news/policy-twenty-tips-for-interpreting-scientific-claims-1.14183

## 1. "Lying" with statistics

i) Conservation biology can be a very controversial discipline and the data produced by researchers in the field are often used to push specific political or social agendas (though generally not by the researchers themselves). Consequently, it is critical that you are able to assess the numbers that you see quoted (in the newspaper, on CNN, on the web, in my lectures) and decide whether you should really believe them.

ii) The main goal of this lecture is to give you a perspective on the way in which even very simple statistics can be misused or misinterpreted. The primary goal is not for you to memorize every example I give (some of which do not come from conservation biology), but to understand the problems with each of them so that you can correctly assess the veracity of any statistics that you come across.

## 2. Problems with numbers

A) BIASED DATA

Things to think about include:

i) Are the data that the relationship is based upon a representative (i.e., unbiased) selection of all the possible individuals that could have been sampled? E.g., think about the paradox of the falling cats.

ii) Could the researchers have verified the accuracy of their data by some independent means? E.g., survey respondents might give false information.

B) LYING WITH AVERAGES AND THE EFFECTS OF EXTREME DATA POINTS

i) When "averages" are reported do you always know what is being discussed?

ii) Note that there are three different types of "average": the mean (arithmetic average), median (middle value in a distribution), and mode (most commonly occurring value).

iii) These three numbers can be very different, and so the "average" value of something may not be typical. In fact, if someone just refers to an "average", you may not be able to tell which of the three numbers they are talking about (often it is the mean – but not always). E.g., the average size of tax breaks that politicians talk about do not necessarily tell you anything about what most people will actually save.

C) SPURIOUS CORRELATIONS

i) Just because two things are correlated with each other (i.e., as one changes, so does the other in some systematic way), you should not assume that one is affecting the other (i.e., that there is causality).

ii) Lots of examples of this. Here is one: people who have an annual physical live for longer than those that do not, yet most tests performed in physicals have no detectable health benefits. Why do you suppose the people who have physicals live longer?

iii) The “Theory of the Stork” is another good example, more directly relevant to conservation biology. We’ll talk about other examples in class.

D) INAPPROPRIATE COMPARISONS

i) It is easy to mislead people by using different types of statistics in a comparison.

ii) E.g., this quote seems to be designed to convince you that dolphins living in captivity die at a higher rate than they would in the wild: "Calculations taken from the study showed that on average the expected life span of a bottle-nosed dolphin in captivity could be as little as 14 years, while in the wild the dolphin could live twenty to twenty-nine years."

iii) Did you spot the problem? Read it again. If you still can’t see it, make sure you are in class. But, for what it is worth, this statement tells us absolutely nothing useful and is apparently intended to mislead you.

E) EXTRAPOLATION BEYOND THE DATA

i) When you look at graphs, pay attention to whether the line through the points goes beyond the range of values spanned by the points. If any assertions are made about values far outside of the data range, you probably should not put too much faith in them (though there may be cases where it is OK to – e.g., if uncertainty is also well described).

ii) This problem is not isolated to graphs (think of the examples given in lecture), but it is on graphs that it is perhaps most misleading.

## 3. Misleading graphics

There are lots of ways to mislead people with clever (or sloppy, depending on the goal) graphics. Here are a few things to watch out for:

i) <u>Axes with no units.</u>

ii) <u>Axes that are nonsense</u> (or not explained).

iii) <u>Tricky visuals</u> (especially those that mess with the size of things in the graphic, in ways that imply that some number is bigger or smaller than it really is).

iv) <u>Broken axes</u> (e.g., that make something look like it is approaching zero – or some other specific value – when it is nowhere near that value).

v) <u>Compressed or expanded axes</u> (especially when two or more graphs are shown side by side in a way that invites comparison).

vi) <u>Playing with the y-axis</u> (especially the use of log scales, which can make changes seem less extreme than they are).

vii) <u>Titles that do not match the large scale pattern.</u>

There are legitimate reasons for doing some of these things – e.g., broken axes or the use of a log scale – but they can still mislead the unwary, and be misused by the devious.