

How to handle zeroes in ecological data

February 16 2016, by Alexandra (Sasha) Wright

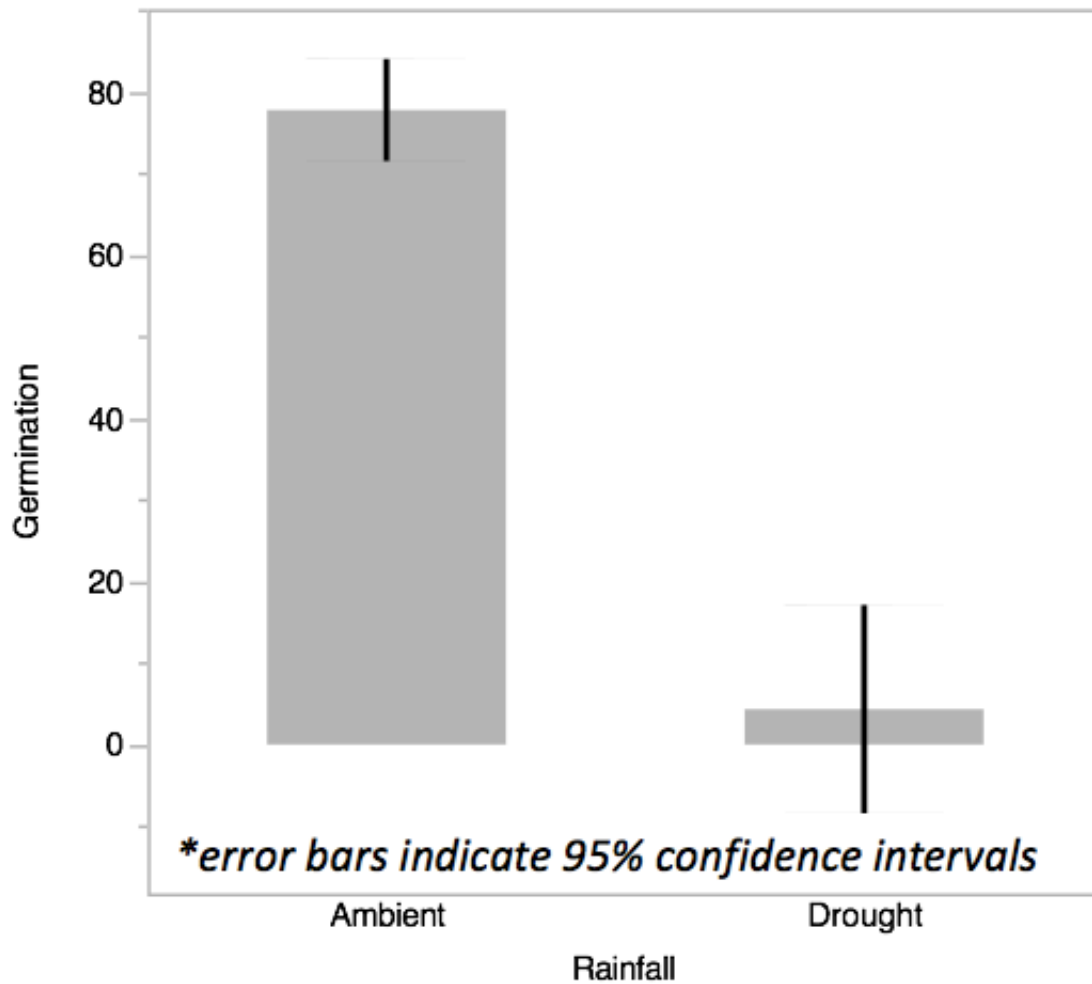


Fig. 1 Seedling germination (%) in response to drought.

The analysis of ecological data can be a difficult endeavor. Ecological data are noisy: some days are windy, some days are hotter than usual, sometimes ants chew through your carefully placed flagging tape, and sometimes your entire experiment disappears overnight. It's an experimental crime scene. We usually deal with these myriad problems with a variety of fancy statistics and massive sample sizes. But even before the monkey sh*t hits the fan, there is an incredibly common data-related question that most plant ecologists face: if I want to record the size of a seedling in my experiment, but the seedling never germinated in the first place, should I record it as a zero or a blank? Some version of this question has come up in the context of every ecological experiment I have ever been involved with. It is glossed over in the vast majority of manuscripts I review. And it has massive implications for our interpretations of data. So how should we be handling the zero/blank question in ecology?

Imagine this scenario: you are interested in assessing the effects of drought on wheat fitness. In order to get a well-rounded understanding of fitness you will likely want to measure [wheat plants](#) at several points during their life cycle: [seed germination](#), annual yield, percent plant survival, flower number, seed number, and germination success of offspring.

To do this experiment you start by planting some number of wheat seeds (let's say 100) into 3 plots with ambient rainfall and into 3 different plots with rainfall that you have reduced by 50%.

Your data may look something like this:

Plot	Rainfall	Germination	Yield (g/m ²)	Seedling Survival	Flower Number	Seed Number	Offspring Germination
1	Ambient	78%	120	95%	10	45	10%
2	Ambient	80%	90	98%	7	38	20%
3	Ambient	75%	100	92%	3	38	(I) 0%
4	Drought	10%	30	5%	3	(F) 0	(J) 0%
5	Drought	3%	15	2%	(D) 0	(G) 0	(K) 0%
6	Drought	(A) 0%	(B) 0	(C) 0%	(E) 0	(H) 0	(L) 0%

The reality is that for (almost) every zero in the above datasheet, there is an alternate case to be made for replacing this 0 with a blank. I have labelled all instances (A-L) above.

For all examples, there are three things we have to ask ourselves:

Is the process itself biologically possible given the starting conditions at this point in time? For example, was seedling survival possible given that no seeds germinated in the first place?

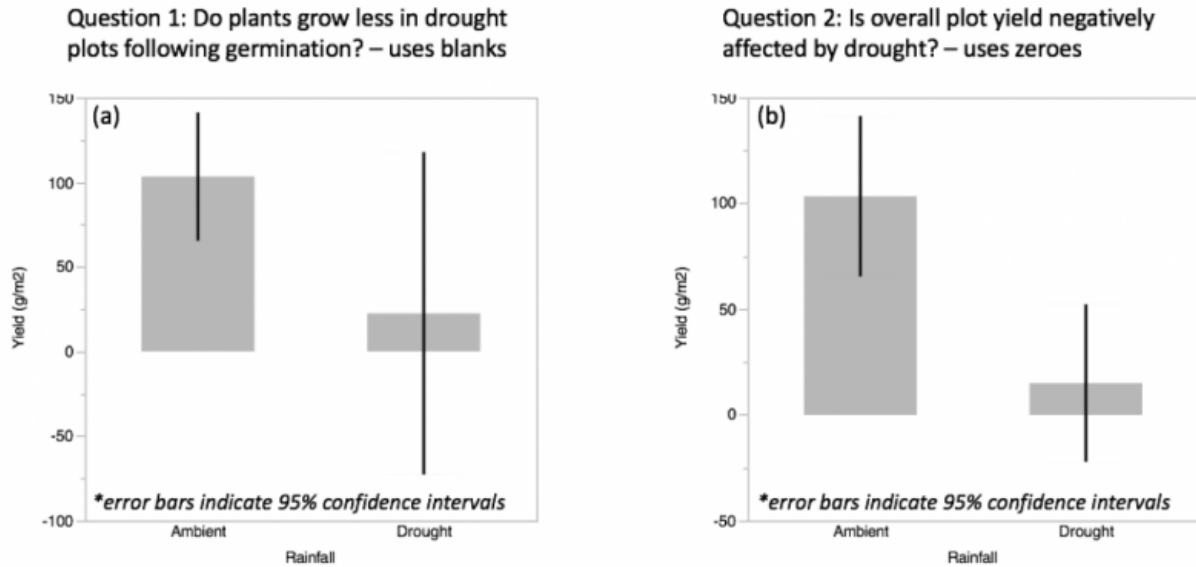


Fig. 2 Plot yield in response to drought considering two potential ecological questions.

What question are we trying to answer?

How will we interpret the data we collect to address this question?

Let's explore each case in sequence:

Case A. This is question about percent germination in the plot. In our hypothetical scenario, we planted 100 seeds and 0 germinated in this plot. If we use the above questions as a guide, the answers are quite straightforward:

(1) Was germination possible in this plot? Yes, 100 seeds could have germinated.

(2) What question are we trying to answer? Does drought affect wheat germination rates?

(3) How do we interpret the data (Fig. 1)? Wheat seeds germinated at a lower rate in drought conditions.

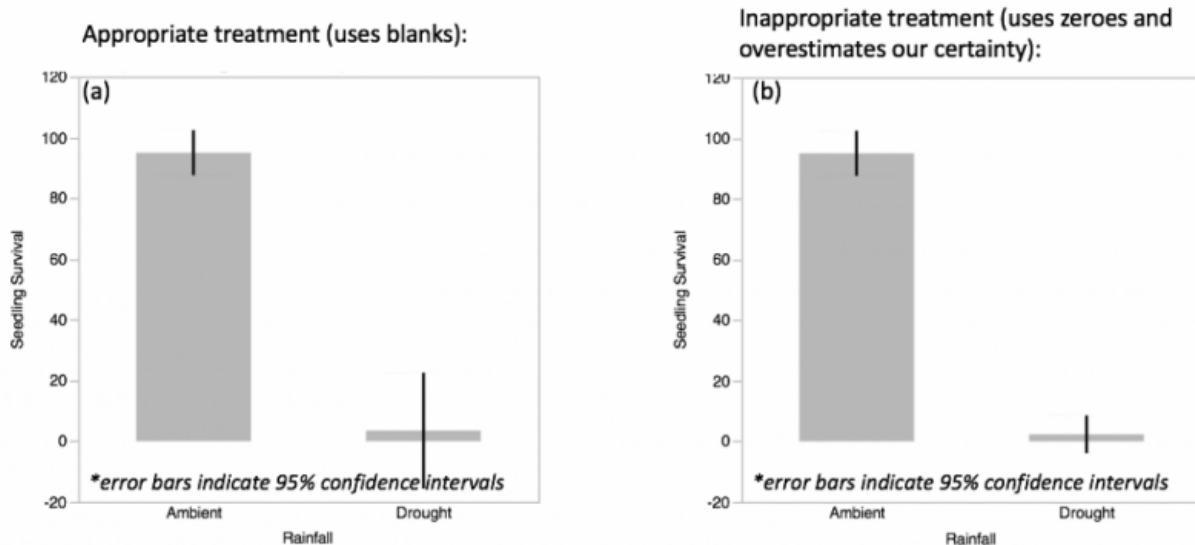


Fig. 3 Seedling survival (%) in response to drought considering an appropriate treatment of zeroes and an inappropriate treatment of zeroes.

Thus, there is nothing that suggests that this zero is inappropriately recorded

Case B. This is a question about overall yield in the plot. Importantly, there are at least two ways to interpret this metric and therefore record the data. Again, we can use the above framework.

(1) Was it possible for the plot to yield biomass? This depends. If we consider that none of our seeds germinated, yield should be impossible. Things obviously can't grow if they didn't germinate in the first place. If the process itself is impossible, we should be leaving this cell blank (or filling it with an NA that will be ignored in any analyses).

On the other hand, it depends on what we mean by yield. In community ecology, yield is often used as a proxy for overall performance of an experimental treatment. We often use yield to integrate over all processes: germination, growth, and survival. If we leave the cell blank, we miss this very important part of the story: sometimes drought plots fail completely. In this case, a 0 is an appropriate value for this cell. But ultimately it depends on the question we are asking.

(2) What question are we trying to answer? There are really two possible questions that could be answered with these data. We could ask: do the plants grow more in ambient conditions than in drought conditions following germination? Or the broader question: Is overall plot yield (performance) negatively affected by drought?

(3) How do we interpret the data in the context of these two separate questions? Question 1: plants do not necessarily grow less in drought plots following germination (overlapping 95% confidence intervals, Fig. 2a). Question 2: plot yield is negatively affected by drought (Fig. 2b). The important point here is that if we used the wrong data (Fig. 2a) to answer the broader question (question 2), we might wrongly conclude that drought does not negatively affect plot yield! An inaccurate conclusion!

Case C. Survival in the plot: here again, there are two possible things we could record, though in this case one is more appropriate than the other.

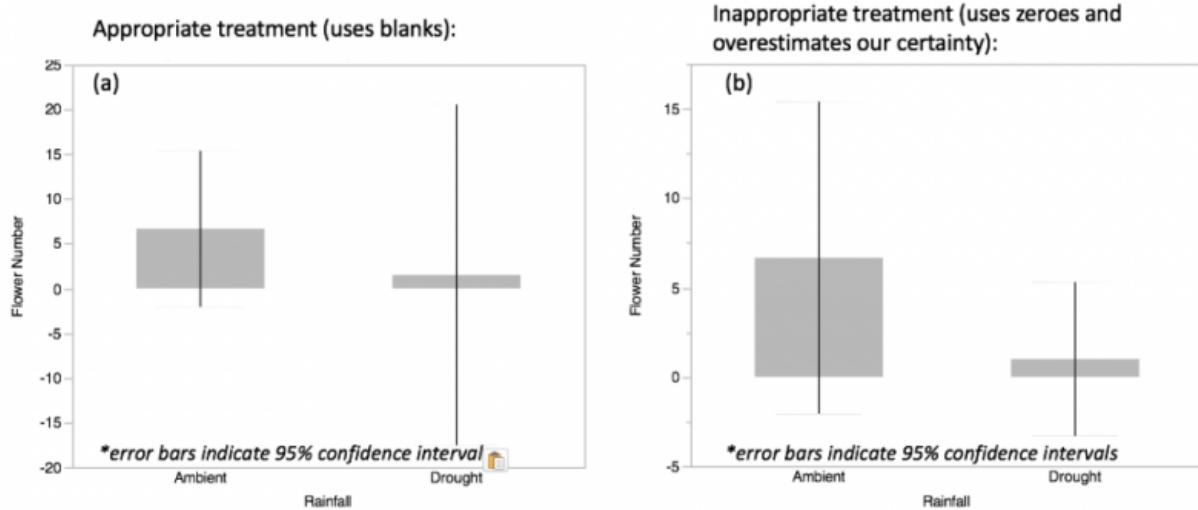


Fig. 4 Plot flower number in response to drought considering an appropriate treatment of zeroes and an inappropriate treatment of zeroes.

(1) Was it possible for these seedlings to survive? Considering that we measured germination, and germination in this plot was 0%, the answer is no. You can't assess whether or not the plant died, if there was no plant growing there in the first place. This cell should be recorded as a blank (or NA).

That said, if we hadn't measured germination, it is often common in ecology to use survival as a proxy for both germination and survival. In that case, we could record a 0 in this cell, but we would need to be careful in our interpretation. We wouldn't know if these were differences in survival or differences in germination rates.

(2) What question are we trying answer? Does drought negatively affect wheat survival?

(3) How do we interpret the data in the context of this question? Drought negatively affects seedling survival (Fig. 3a). If we were to include the 0 in our analysis we would incorrectly overestimate our certainty of this effect (Fig. 3b).

Case D. Flower Number: This response is correctly recorded as a 0 according to our criteria.

(1) Was it possible for these plants to grow flowers? Yes. There were plants growing here, so it was possible for these plants to allocate some of their energy to reproduction.

(2) What question are we trying to answer? Does drought negatively affect flower production?

(3) How do we interpret the data in the context of this question? See below.

Case E. Flower Number: This response was incorrectly recorded as a 0 (should be NA).

(1) Was it possible for these plants to grow flowers? No. There were no plants left in these plots. It was therefore impossible to measure whether or not plants allocated some of their energy towards flower production. This should be recorded as a blank.

(2) What question are we trying to answer? Does drought negatively affect flower production?

(3) How do we interpret the data in the context of this question? We don't have enough data to conclude that plants invested less in flowers in drought conditions (Fig. 4a). Incorrectly including this data point as a zero gives the impression that we are more certain of the differences

than we actually are, but is still insufficient to conclude anything about differences (Fig. 4b).

Cases F-L follow the same logical reasoning as particular cases above. See Appendix at bottom for details.

The corrected data table (according to the above arguments should be):

Plot	Rainfall	Germination	Yield (g/m ²)	Seedling Survival	Flower Number	Seed Number	Offspring Germination
1	Ambient	78%	120	95%	10	45	10%
2	Ambient	80%	90	98%	7	38	20%
3	Ambient	75%	100	92%	3	38	0%
4	Drought	10%	30	5%	3	0	NA
5	Drought	3%	15	2%	0	NA	NA
6	Drought	0%	0	NA	NA	NA	NA

Based upon the above, consider these guidelines:

In data that are collected sequentially (e.g. flower number is recorded and then seed number is recorded at a later date), there should be necessary contingencies built into your data table. If the first event (flower number) didn't happen, then the second event (seed number) is impossible. This should be recorded as a blank.

If the variable being measured (e.g. yield) is usually interpreted as the integrated response of many different biological processes (e.g. [germination](#), growth, and survival), you should record a 0 based on whether ANY of those responses were possible.

In community ecology, both plot yield and survival are often used as catch-all metrics: many biological processes are assumed to contribute to the number that is recorded. Be sure you are correctly interpreting what these catch-all metrics in your data actually mean.

More information: Appendix:

Case F. Seed Number: This response is correctly recorded as a 0 for the same reasons as Case D.

Cases G & H. Seed Number: if there are no flowers, there is no opportunity to know whether the plant would have produced seeds. These responses are incorrectly recorded as zeros for the same reasons as Case E.

Case I. Offspring Germination: This response is correctly recorded as a 0 for the same reasons as Cases D & F.

Cases J-L. Offspring Germination: if there were no seeds produced, there is no potential for germination. In fact, a germination trial couldn't have been run. These responses are incorrectly recorded as zeros for the same reasons as Cases E, G, and H.

This story is republished courtesy of PLOS Blogs: blogs.plos.org.

Provided by Public Library of Science

Citation: How to handle zeroes in ecological data (2016, February 16) retrieved 20 July 2024 from <https://phys.org/news/2016-02-zeroes-ecological.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.