

Null model power for multiple stressors

David Murrell

23/08/2022

Power analysis for stressor interactions

This tutorial guides users through the code provided by Burgess, Jackson and Murrell (2022) to estimate the power to detect a departure from the null model for stressor interactions. The code implements either the additive or multiplicative null models.

When to use the power function

Power analyses should always be performed **before** an experiment, and acts as a guide to the limits of what you can expect to detect given your experimental design (ie number of replicates).

How should the numbers for the function be chosen?

It is best to choose *plausible* numbers (i.e. biological responses to no stressors, to each stressor acting independently, and to both stressors co-occurring) that might be based on previous or preliminary results.

Getting started

First we need to load the R library `multiplestressR`. You may need to install this from CRAN.

```
library(multiplestressR)
```

You need to get hold of the function `interaction_power` using the R code “`interaction_power.R`”. Make sure you are pointing your computer to the directory where the code is stored. In the example below the R code has been stored in my Downloads folder.

```
source("~/Downloads/interaction_power.R")
```

An example using the additive null model

Let us assume we are considering density of a population as our response variable, and that we expect our two stressors to have a negative effect on density. We make the following assumptions:

1. Our control treatment has a mean density of 80.
2. The mean population density under each stressor acting alone are depressed to 65, and 65 respectively.
3. The mean population density when both stressors are occurring is depressed (relative to the control) to 60.
4. The standard deviations around the mean treatment values are all equal to 5.
5. We have 4 replicates for each treatment.

We also estimate power over 1000 simulations, and we use the additive null model “A” option. The function assumes a 5% level of significance is being used.

What are we testing?

We are testing if our observed response to the interaction can be distinguished from the null model expectation that the stressor effects simply add up.

What does the function do?

The function helps us understand if our experimental design would allow us to detect a non-additive interaction between our two stressors.

The code to implement this is

```
interaction_power(80, 65, 65, 60, 5, 5, 5, 5, 4, 4, 4, 4, sims=1000, model="A")
```

```
## Expected interaction treatment: 50
```

```
## power mean_effect_size detected_ES
## 1 0.352          -1.681432    -2.698978
```

Note the output. The first line tells us the expected population density under the additive null model is 50. It is good to check this as the additive model can lead to negative numbers, or if using rates (e.g survival), expected rates that are greater than 1. This also tells us that our interaction is antagonistic since the assumed observed interaction response (60) is greater than the expected interaction response.

The second set of numbers relate to power, mean effect size, and mean detected effect size. We deal with these in turn.

1. Power. In our example, power is 0.352, meaning our experimental design would detect the interaction as being significantly different to the null expectation in 35.2% of experiments. This is quite low, and remember this is an estimate based on 1000 simulations. If we repeated the simulations the random number generator would lead to slightly different numbers coming out, but as long as we use a reasonably large number of simulations (at least 1000), there shouldn't be too much variation and our estimate is in the ball-park.
2. Mean effect size. This is the mean of the effect sizes for the interaction across all simulations. Since it is negative, we can confirm the interaction is antagonistic.
3. Mean detected effect size. We now take the mean of the experiments that return a significant result. We see that the mean is larger than the mean effect size. This shows that when we have low power we can greatly overestimate the strength of the interaction.

Summary

Under our current experimental design we

1. Have low power to detect our interaction as being non-additive even though the treatment means suggest it might be antagonistic.
2. If we proceed with our design, and we get a statistically significant difference to the null model, then there is a high chance this will be an outlier, and we are overestimating the strength of non-additivity. This is because we can only detect the extremes of the distribution of 'true' values.

So, perhaps we are able to increase the number of replicates for each treatment to 20. What is the effect of power?

```
interaction_power(80, 65, 65, 60, 5, 5, 5, 5, 20, 20, 20, 20, sims=1000, model="A")
```

```
## Expected interaction treatment: 50
```

```
## power mean_effect_size detected_ES  
## 1 0.993 -1.966917 -1.974983
```

Note we have very high power (>99%), and now we are not overestimating the strength of the interaction when we do detect a non-additive response to the stressors co-occurring.

Exploring different scenarios

We could also change some of our values. For example, what if the response to both stressors is closer to the expectation? Let's assume the mean density under both stressors is now reduced to 55:

```
interaction_power(80, 65, 65, 55, 5, 5, 5, 5, 20, 20, 20, 20, sims=1000, model="A")
```

```
## Expected interaction treatment: 50
```

```
## power mean_effect_size detected_ES  
## 1 0.595 -0.9913802 -1.278109
```

We now find the power has dropped. Not surprising, but we are perhaps reaching the limits of what we could reliably detect given our replicate numbers. Perhaps this is ok because the interaction response is getting to be similar to the expectation, and is close enough to additive to be not worth detecting. This is something for the researcher to determine.

An example using the multiplicative null model

We could instead assume the interaction is multiplicative and use the same numbers on the multiplicative null model. The code looks very similar:

```
interaction_power(80, 65, 65, 60, 5, 5, 5, 5, 4, 4, 4, 4, sims=1000, model="M")
```

```
## power mean_effect_size detected_ES  
## 1 0.426 -0.1301834 -0.19399
```

Note we get different numbers! This is not surprising, and don't be surprised if the effect sizes sometime change sign. However, the interpretation of the numbers are the same as before. In this case the effect sizes are negative meaning the interaction is still antagonistic.

The bottom line is the null models are different and shouldn't be compared. *Indeed, we should decide up front which model is more appropriate for our data and not cherry pick based on power etc.*

References

- Burgess, B.J., Jackson, M.C. and Murrell, D.J., 2022. Are experiment sample sizes adequate to detect biologically important interactions between multiple stressors? *Ecology and Evolution* (accepted)

The following article describes the `multiplestressR` package

- Burgess, B.J. and Murrell, D.J. (2022) `multiplestressR`: An R package to analyse factorial multiple stressor data using the additive and multiplicative null models. *bioRxiv* 2022.04.08.487622; doi: <https://doi.org/10.1101/2022.04.08.487622>