TIEE

Teaching Issues and Experiments in Ecology - Volume 20, June 2024

Reorganisation following disturbance: multi trait-based methods in R

Part 3: Analysing the trait-based diversity indices and presenting your data

Estimated time: 120 minutes

In Part 2, you ran a PCoA to visualize the functional trait space and ran code to compute the multi-trait-based diversity metrics pre and post bleaching based on species coordinates in the trait space. You also produced summary graphs comparing multi-trait-based diversity (functional richness), species richness, and total abundance before and after bleaching in your habitat of focus.

Today, we will do two things:

1. Test for differences in the fish assemblage metrics before and after bleaching
2. Review what makes for an effective scientific poster

Testing for differences between paired means

Aim: apply a t-test to test for differences between the fish metrics before and after bleaching.

Note: we will go through one example using the estimates of multi-trait-based diversity but you can amend and apply this code to compare the abundance estimates for your species of focus (Part 1), species richness and total abundance (Part 2) – just make sure the format of your data-file matches the example below.

Which statistical test you use depends on your research question and the type of data you have. Having a clear idea on your question and your data types will make identifying the correct statistical tool much easier.

Here, the question is whether bleaching (a two-level categorical explanatory factor - before or after bleaching) has an effect on a multi-trait-based diversity (a continuous response variable).

Comparing two means — two samples

So in data type terms, we are interested in the effect of a two-level categorical variable on a continuous quantitative response variable. There are three types of t-test, each appropriate for a different situation of comparing two mean estimates. For convenience in R, the single `t.test()` function be applied to situations in which you may want to use a t-test:

A one sample t-test Use to compare a single sample mean against a hypothesised value for the population mean.

A two sample t-test Use to compare the means of two groups, and test the null hypothesis that they have the same mean. It comes in two forms: the Student test assumes that the groups have the same standard deviation, the Welch test does not.

A paired samples t-test Use when you have two measures taken from each individual / group / population, and you want to test the null hypothesis that the two measures have the same mean e.g. weight of fish before and after a diet treatment. It is equivalent to taking the difference between the two measures for each individual / group / population, and then running a one sample t-test on the difference scores.

Paired data are generally less common (although can be very important) than unpaired data. Paired data can be tested using a paired t test or, if the data are non-parametric, then a Wilcoxon signed ranks test.

Q1 Given the three scenarios described above in which you can apply a t-test, which one applies to dataset you are working with?

The t-test

In a nutshell: a t-test involves three steps:

1. calculating the difference between two means
2. dividing by the standard error of the difference
3. comparing the statistic with the value of the students t from a distribution table.

Means are said to be significantly different when the calculated value of the t-statistic is larger than a critical test statistic value (which is derived from the t-distribution).

Remember – the research question is whether bleaching affects multi-trait-based diversity. We will apply a t-test to the data, to determine the answer if the means estimates of multi-trait-based diversity are the same pre and post bleaching.

Q2 Given the research question and the test, write down the null and alternative hypothesis for how multi-trait-based diversity might differ before and after bleaching.

NOTE: we typically don't report the null hypothesis in our final reports or summaries of our research and you shouldn't include them in the poster, but the alternative hypothesis and / or a rewording of it that states our prediction of what will happen can be useful to place at the end of the introduction.

For this reason, you should write down your null and alternative hypotheses for how the abundance of your two species of interest may change, and if you plan on presenting it, how total abundance and species richness will change. The alternative hypothesis should be informed by your reading into the ecology of your species (Part 1), and how it specifically might be impacted by the change in habitat bleaching.

Let's first make sure you have the dataframe of the functional richness estimates loaded into R.

Open the R project that you have been using during these practicals, and start a new script called Day 3. You should find the object you created in Part 2 with all the data for your habitats (i.e., functional richness, fish abundance, species richness etc.) in your R project environment. Alternatively, if you saved the object at the end of Part 2 as a csv file for practice in saving and later loading, you can run one of these following lines of code to load in the dataset. Run the line of code that corresponds to the habitat type you have been assigned to work with: FD_BP = branching Porites, FD_LC = low coral cover, FD_MC = mixed coral, FD _SC = soft coral

```
FD_BP <- read.csv("FD_BP.csv")
FD_LC <- read.csv("FD_LC.csv")
FD_MC <- read.csv("FD_MC.csv")
FD_SC <- read.csv("FD_SC.csv")
```

There are three data columns of interest in this dataframe:

Abun_tot = fish abundance (all species) sp_richn = species richness (number of species) fric = estimate of functional richness

Checking test assumptions before testing for differences between the estimates of multi-trait-based diversity

Before using a t-test to assess for differences between the estimates of multi-trait-based diversity and then writing up your results, you need to first check whether the data meets the assumptions of the test. If they don't, we may have to use the non-parametric equivalent of the t-test.

```
# Test for equal variances
var.test(data=FD_BP, fric~Bleaching) ## p > 0.05 = equal variances

# Test for normaly distributed data
shapiro.test(with(FD_BP, fric[Bleaching == "Pre"] - fric[Bleaching == "Post"]
)) ## p > 0.05 = normally distributed
```

```
        F test to compare two variances

data:  pre and post
F = 0.67558, num df = 17, denom df = 17, p-value = 0.4271
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2527132 1.8060237
sample estimates:
ratio of variances
         0.6755783


        Shapiro-Wilk normality test

data:  pre - post
W = 0.79255, p-value = 0.001194
```

Output from a F-test and Shapiro test run in R.

Parametric tests generally require equal variances and normally distributed data. Here we use the Fisher F test and the Shapiro test to test these assumptions of the data, and so determine whether we are ok to proceed with the parametric t-test to compare means, or whether we might have to use the non-parametric equivalent.

Fisher's F test is also referred to as the F test for equal variances. The F test calculates the ratio of variances of different groups. This ratio is compared to a critical value, which depends on the degrees of freedom (which in turn is dependent on the sample size), and the size of alpha (which is by convention set at 0.05). When the ratio of variances is greater than the critical value, the null hypothesis (H0 = variances are equal) is rejected i.e. the variances are not equal and so you should not proceed with the parametric test.

The function in R that runs Fisher's F test is `var.test(data1,data2)` where data1 and data2 are the vectors containing data from the two samples, or in this case Pre bleaching fric versus Post bleaching fric in branching Porities habitat. The F value is the ratio of variances. If the p-value is greater than the cut-off of 0.05, we accept our null hypothesis.

Note: the Fisher's F test can only be used when comparing the variance of two groups. When testing or comparing two or more groups, you can use the Levene's test.

In this instance, the null hypothesis is that there is no difference between the variances of the two groups, so we can proceed with the parametric test. This can be confusing, because typically we look for a $p < 0.05$. This is because we are looking for there to be differences in what we are testing. In this instance, of checking for homogeneity of

variance, a non-significant result means the variances are the same. This means we can proceed with the parametric test.

The Shapiro test assesses data from a sample with the null hypothesis that the dataset is normally distributed, i.e. the range of data values are distributed equally around the mean value, with data near the mean being more frequent in occurrence than data far from the mean. When graphed, the normal distribution looks like a bell-shaped curve. If the p-value from a Shapiro test is < 0.05, we can reject the null hypothesis and say that the sample was not generated from a normal distribution.

The function in R that runs Shapiro's test is 'shapiro.test(data1-data2)'. The W value is a measure of the difference between a normal distribution and the distribution of the data observations. The W value will be between 0 and 1, with 1 being a perfect match.

Based on the outputs of these tests, decide whether the t-test is the most appropriate test to use, or whether you should use the non-parametric equivalent test, the Wilcoxon signed rank test (code for this below).

Note: the Wilcoxon signed rank test can be applied to paired data.

```
wilcox.test(data=FD_BP, fric~Bleaching, paired=TRUE, alternative = "two.sided
")
```

Testing for differences between the estimates of multi-trait-based diversity (functional richness)

If your data meets the test assumptions so that you can use a parametric test, running a t-test is R is pretty easy.

The following code will, for branching Porites, run a t-test on the functional richness estimates pre and post bleaching. Edit this code to use your habitat type dataframe, add it to your script and pass to R console:

```
# Example using branching Porites habitat and functional richness

t.test(data = FD_BP, fric ~ Bleaching, paired=TRUE) # run the t-test
```

You will see an output from the t test similar to this:

```
        Paired t-test

data:  pre and post
t = 0.031732, df = 17, p-value = 0.9751
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -0.03349405  0.03451694
sample estimates:
mean difference
   0.0005114451
```

Output from a t-test run in R.

Here, the p-value is > 0.05 , therefore we accept the null hypothesis, there is no difference between the mean estimates of functional richness before and after bleaching.

These results can be summarized up as:

Multi-trait-based diversity (functional richness) of the fish community on branching Porites habitat was not different after bleaching (t = 0.032, df = 17, p-value = 0.975).

This is standard format for presenting the results of a t-test – use it for your posters.

Q3 Does fish species richness, functional richness, and total abundance differ before and after bleaching for your habitat?

Using the code provided above, edit and test for differences in species richness (sp_richn), total abundance (Abun_tot), and functional richness (fric) pre and post bleaching in your habitat. Check the test assumptions, and if they are not met (results of the var.test and shapiro.test < 0.05) use the Wilcoxon test instead. Write a summary sentence for each fish metric tested. Cross check the statistical results with the graphs you produced last week – do they match?

Habitat group _____

Multi-trait-based diversity (functional richness):

Total abundance:

Species richness:

In your own time, you should also use the code provided above, to edit and test for differences in your two species of focus.

Moving Forwards – Pulling your results together for your poster

Ok – well done – you now have summary estimates, statistics, and graphs – these will form the results section of your poster.

We will now discuss what makes for an effective scientific poster (see 'Poster_preparation_guide.pptx'). Use the rest of the time in the practical session to ensure you have the results you want to present, and get your R questions answered.

For the poster assignment you should:

1. Listen to the lecture on instructions for making your poster and start drafting bullet points that you will want to make in Background and Data analysis sections.
2. Read the marking criteria for the poster, so you understand what is expected of you.
3. Draft the text for your poster, create your poster in PowerPoint (or some equivalent) and prepare your poster presentation talk.
4. Watch this video for tips on poster making: https://www.youtube.com/watch?v=AwMFhyH7_5g

See you at the scientific poster session!