

## Practical Stats Newsletter for October 2017

Subscribe and unsubscribe: <http://practicalstats.com/news>

Archive of past newsletters <http://practicalstats.com/news/archive.html>

In this newsletter:

- A. Practical Stats Courses
- B. Multivariate Permutation Tests
- C. Webinar: Stats for Managers

### A. Practical Stats Courses

Our Applied Environmental Statistics courses are on our online training site:

<http://practicalstats.teachable.com/>

The two courses separately are each \$650 for a 1-year access for one person. Or get both courses together in a bundle for \$1200.

Next up is *Untangling Multivariate Relationships*, our course on applying multivariate methods to chemical and community structure data, including the newer permutation alternatives (see below). It will be available by the end of December on our training site.

### B. Multivariate Permutation Tests

Looking at chemical or species variables one by one often misses the point – there is interaction between variables, and the entire pattern holds valuable information that looking at each individually will not capture. Traditional multivariate methods such as Multivariate Analysis of Variance (MANOVA) require data to have multivariate normality, a more strict condition than just that each of the variables individually follow a normal distribution. MANOVA extends ANOVA group tests to the case of multivariate response (Y) variables – does the pattern of multiple variables differ between groups? Yet like ANOVA, MANOVA may not work well in the common situation when data are not multivariate normal. What alternatives are there to discern differences between groups?

As an example, physical/chemical conditions and co-located benthic macrofauna abundance were measured in a radial pattern around an oil-drilling platform (Warwick and Clark, 1991). Four zones of distance out from the platform were defined, and differences in macrofauna going away from the platform were examined. The nine physical/chemical characteristics used as correlative information included total hydrocarbons, mean substrate particle size, and six trace elements including barium (Ba) and strontium (Sr). Does the pattern of these nine characteristics differ between the four zones?

First, Mardia's multivariate test of normality was conducted on the collection of the nine variables (conclusion – not multivariate normal). Figure 1 shows the multivariate Q-Q plot using the nine variables. If data were multivariate normal their pattern of dots would closely follow the solid line. It doesn't.

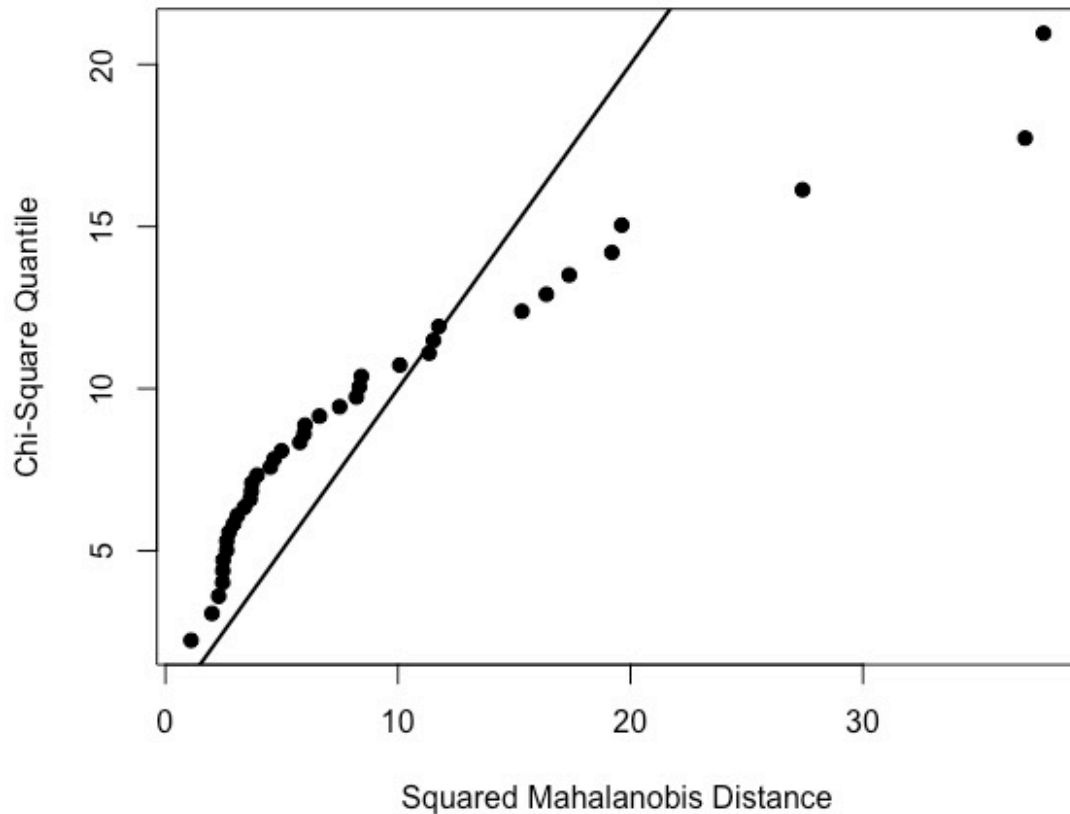


Figure 1. Multivariate Q-Q plot of the pattern of nine physical/chemical variables

This calls into question the validity of a MANOVA test. If variables are skewed, using a mean to represent the middle also is questionable. Two alternative tests are available. First, the permutation equivalent of MANOVA is the MRPP (Multi-Response Permutation Procedure) test. MRPP uses a permutation process to compute a distribution-free test of whether multivariate means of the four groups differ. This answers the issue of non-normality, though it still leaves the mean as the appropriate measure of center. The R command for MRPP is:

```
> mrpp(dat, DistGp, permutations = 9999)
```

```
Dissimilarity index: euclidean
Weights for groups: n
```

```
Class means and counts:
```

	1	2	3	4
delta	1308	846.5	715	636.2
n	6	10	12	11

```
Significance of delta: 0.0001
Permutation: free
```

The small p-value shows that the four group patterns do differ.

A second test is the nonparametric ANOSIM test (Clarke, 1993). ANOSIM also computes p-values by permutations, comparing nonparametric correlations between the pattern of group assignments and the pattern of similarities between measurements of the nine variables. The test determines whether the multivariate percentiles differ between groups – do some variables have higher values in one group over another? ANOSIM does not require multivariate normality. It also looks for 'typical' differences rather than using the mean as its measure of the center, and is therefore more appropriate when variables are skewed. R performs this by:

```
> anosim(dat = eko.std, grouping = DistGp, distance = "euclidean",  
         permutations = 9999)
```

```
ANOSIM statistic R: 0.3429  
  Significance: 0.0001  
Permutation: free
```

Again, the p-value is small, indicating that at least some of these chemical/physical parameters, or their inter-relationships, differ between the groups. Given that environmental data in air, soils, water, and counts of organisms rarely follow a normal distribution, these two methods are extremely useful for determining whether characteristics differ between groups.

Extensions to ANOSIM also determine which of the variables best distinguish among the groups. They found that using only two variables, Ba and Sr concentrations in the sediments, differentiate the four groups of locations better than all nine variables do. Being able to determine which variables most indicate differences leads to a better explanation of why the groups differ. These two permutation tests and much more are presented in our Untangling Multivariate Relationships course. Look for it soon in our Online Training Center.

#### *References:*

Clarke, K. R., 1993. Australian Journal of Ecology v.18, p. 117-143.

Warwick, R.M. and K.R. Clark, 1991. Journ. Marine Biol Assoc. UK, v. 50, p. 225-244.

#### C. Webinar: Stats for Managers

*Back by popular demand – Stats for Managers, a 2015 webinar.* This free webinar covers new methods in statistics that managers of technical staff should know exist – new types of tests, free and powerful software, etc. The webinar doesn't go through the mechanics of the methods, only their importance for doing data analysis in today's world. There are a ton of new methods beyond t-tests – make sure your employees are up to date. Look for the 1-hour webinar at:

<http://practicalstats.teachable.com/>

'Til next time,

Practical Stats

-- Make sense of your data