Practical Stats Newsletter for December 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 - Permanova
C. Course Outline for Untangling Multivariate Relationships

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 soon on our training site.

B. Multivariate Permutation Tests - Permanova
As noted in October's newsletter, 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 follows a normal distribution. What alternatives exist that do not require multivariate normality to discern differences between groups?

As previously presented in October's newsletter, 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). The nine variables do not form a multivariate normal distribution. Still, we want to know whether the pattern of these nine characteristics differs between the four zones.

In October we presented the MRPP (Multi-Response Permutation Procedure) and ANOSIM tests. MRPP uses a permutation process to compute a test of whether multivariate means of the four groups differ. It computes metric, linear dissimilarities between groups analogous to multivariate regression, so that it is not a 'nonparametric' test in the sense of a rank-based procedure as is

ANOSIM. It differs from Permanova, this month's test, in that the process doesn't compare within and between group sums of squares - it is not directly analogous to MANOVA.

We also presented 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 can be considered a multivariate expansion of the Kruskal-Wallis test. It looks for 'typical' differences rather than using the mean as its measure of the center, and is therefore more appropriate when variables are skewed. But what if we actually want a test between groups for differences in the means of multiple characteristics?

Permutation MANOVA or Permanova (Anderson, 2001) extends the concept of permutation approaches from univariate Analysis of Variance. The F statistic for parametric MANOVA is computed, but rather than assuming a multivariate normal distribution and equal variance/covariance matrices for each group in order to compute the p-value for the test as in MANOVA, group assignments are permuted thousands of times and the test statistic stored for each permutation. The p-value is then the proportion of permuted test statistics that are equal to or more unusual than (greater than) the one observed test statistic from the original data. This remains a test for whether the means of multivariate characteristics differ between groups. Its computation is directly analogous to MANOVA in that it computes an F test from between and within group variations in data.

For the Warwick and Clark physical/chemical data used in October's newsletter, the Permanova test is computed in R using:

```
> adonis(formula = eko.9Xvars ~ DistGp, permutations = 9999, method =
"euclidean")

Permutation: free
Number of permutations: 9999

Terms added sequentially (first to last)

          Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
DistGp     1  20892428 20892428  30.178 0.44922  1e-04 ***
Residuals 37  25615470   692310         0.55078
Total     38  46507898                  1.00000
```
and is again highly significant.

Permanova is the preferred approach to differentiate groups based on multiple variables if the mean is the important characteristic to be measured. ANOSIM is

the preferred approach when a truly nonparametric test of differences in percentiles (cumulative distribution functions) between groups is desired, as a multivariate expansion of the Kruskal-Wallis test. 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:
1. Clarke, K. R.,1993. Australian Journal of Ecology v.18, p. 117-143.
2. Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. Australian Ecology v26, p.32-46.
3. Warwick, R.M. and K.R. Clark, 1991. Journ. Marine Biol Assoc. UK, v. 50, p. 225-244.

C. Course Outline for Untangling Multivariate Relationships
Here are the topics covered by our multivariate analysis course, coming soon to our Online Training Center http://practicalstats.teachable.com/ :

1. Introduction to Multivariate Methods - how these methods work
2. Graphing Multivariate Data - views of data in multidimensional space
3. Principal Component Analysis -- looking from a fresh perspective
4. Factor Analysis -- what lies beneath?
5. Cluster Analysis -- grouping ungrouped data
6. Nonmetric Multidimensional Scaling -- a nonparametric ordination method
7. Correspondance Analysis and Principal Coordinates Analysis
8. Relating species patterns to environmental variables -- Mantel tests, PLS, and canonical correlation
9. Discerning and testing differences between groups -- MANOVA vs ANOSIM vs MRPP vs Permanova. Also, DFA versus Logistic Regression
10. Multivariate Trend Analysis - Seriation, a "Mann-Kendall" trend test of the pattern of multiple Y variables

'Til next time,

Practical Stats
  -- Make sense of your data