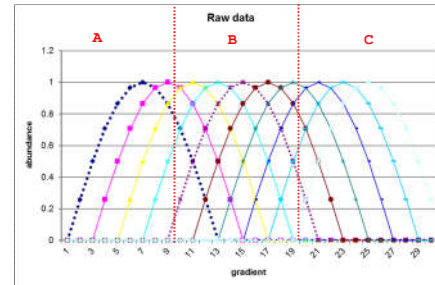


### Indicator Species Analysis

- **What it does:** Analysis aims to identify what species are "indicators" of groups of samples or experimental treatment.
- Typical application involves a priori grouping of samples.



### What is an indicator species?

- **Indicator species** – a species that is characteristic of a group of samples.
  - A perfect indicator species will only occur in one group.
  - Generalist species will occur across multiple groups and therefore not a good indicator.
  - Rare species are not sampled often, also not a good indicator.
  - Indicator values range from 0 to 1 but may also be expressed as a percentage.
  - Each species is assigned an indicator value for each group.
  - Significance of values assessed through permutations.

### Indicator Species Analysis

- Appropriate questions
  - A priori grouping of samples
    - Cluster analysis → find indicators for clusters
  - Disturbance or other discrete environmental variable → what species are indicators for pre/post disturbance?
  - Distinct habitat types → what species are indicators of each habitat?
  - Diet analysis → what diet items are indicators for a species or group from a particular habitat?

### Indicator Species Analysis

- Indicator Values

- Calculated for each species-group combination

- $X_{kj}$ =mean abundance of species j in group k
- $n_k$ =number of samples in group k
- $a_{ijk}$ =abundance of species j in sample i of group k

$$X_{kj} = \frac{\sum_{i=1}^{n_k} a_{ijk}}{n_k}$$

- $RA_{kj}$ =relative abundance of species j in group k

$$RA_{kj} = \frac{x_{kj}}{\sum_{k=1}^g x_{kj}}$$

### Indicator Species Analysis

- $RF_{kj}$ = Relative frequency of species j in group k
- $B_{ijk}$ =presence/absence of species j in sample i of group k

$$RF_{kj} = \frac{\sum_{i=1}^{n_k} b_{ijk}}{n_k}$$

- Indicator Value** – product of the relative frequency and relative abundance.

- A measure of exclusiveness for a species in a group.
- Typically, only the highest indicator value for a species across groups is used.

$$IV_{kj} = 100(RA_{kj} * RF_{kj})$$

### Significance of Indicator Values - Monte Carlo Tests

- Test the significance of indicator values through permutation test
- randomly assign samples to groups, generate a distribution of indicator values.
- Null – observed IV will not differ from those generated through permutations.

### Indicator Species Analysis

- Some other uses
  - Determine stopping point for clustering
    - If the goal is to have clusters that describe distinct communities, groups should contain indicator species
  - One could define optimal clustering as when you maximize indicator values
- Has a few things in common with Polar (Bray-Curtis) ordination. Can be thought of as ordinating species along a categorical environmental gradient or among experimental treatments.

### Indicator species analysis functions

- `indval` function in `labdsv` package
  - Provide community matrix and grouping variable
- `multipatt` function in `indicspecies` package
  - Provide community matrix and grouping variable
  - To do the traditional indicator species analysis
    - Options: `duleg=TRUE` and `func="IndVal.g"`

### Indicator Species Analysis

- Input – raw (not transformed) community data and grouping variable. Analysis assumes count data.
- Code
  - `Indicator_species<-multipatt(community,clust,duleg=TRUE,func="IndVal.g")`
- Technique originally described in:
  - Dufrene, M. and P. Legendre. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345-366.

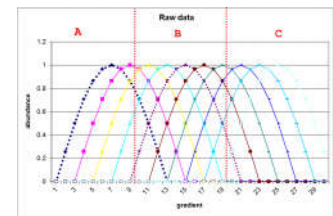
### Output

- Relative frequency of occurrence for each species in each cluster ( $RF_{jk}$ )
- Relative abundance of each species in each cluster ( $RA_{jk}$ )
- Indicator values 100 ( $RF_{jk} * RA_{jk}$ )
  - Note that `multipatt` function returns the square roots, square them if you are reporting them as indicator values.
- Significance of indicator values from permutations
- Group that each species is associated with (highest value)

```
ind_species<-multipatt(community,envdata$spatial)
```

```
$B
      a      b      c
sp1 1.0000000 0.2 0.0
sp2 0.7777778 0.4 0.0
sp3 0.5555556 0.6 0.0
sp4 0.3333333 0.8 0.0
sp5 0.1111111 1.0 0.0
sp6 0.0000000 0.9 0.2
sp7 0.0000000 0.7 0.4
sp8 0.0000000 0.5 0.6
sp9 0.0000000 0.3 0.8
sp10 0.0000000 0.1 1.0

$A
      a      b      c
sp1 0.90918235 0.09081765 0.00000000
sp2 0.71494203 0.28505797 0.00000000
sp3 0.46021386 0.53978614 0.00000000
sp4 0.20993468 0.79006532 0.00000000
sp5 0.03771739 0.96228261 0.00000000
sp6 0.00000000 0.9009958 0.09990042
sp7 0.00000000 0.69299280 0.30700720
sp8 0.00000000 0.43417375 0.56582625
sp9 0.00000000 0.19299280 0.80700720
sp10 0.00000000 0.03407417 0.96592583
```



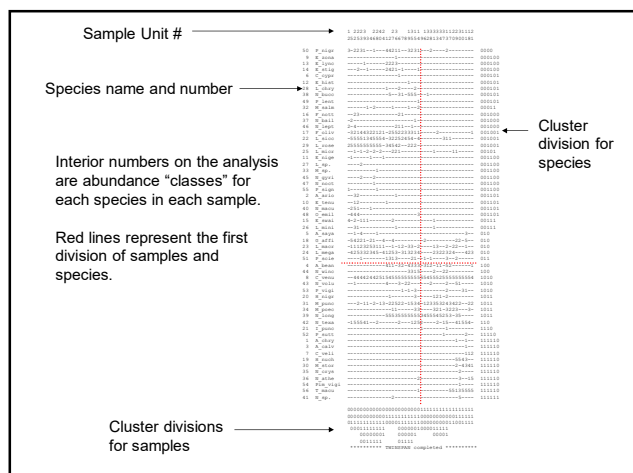
```
$sign
      s.a s.b s.c index      stat      p.value
sp1  1  0  0  1 0.9535105 0.00019996
sp2  1  0  0  1 0.7456983 0.00319936
sp3  0  1  0  2 0.5690973 0.13957209
sp4  0  1  0  2 0.7950171 0.00079984
sp5  0  1  0  2 0.9809600 0.00019996
sp6  0  1  0  2 0.9000498 0.00019996
sp7  0  1  0  2 0.6964876 0.00859828
sp8  0  0  1  3 0.5826626 0.11377724
sp9  0  0  1  3 0.8034960 0.00119976
sp10 0  0  1  3 0.9828153 0.00019996
```

### Two Way Indicator Species (TWINSpan)

- Hill, M. O. 1979. TWINSpan—A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics. Cornell University Press, Ithaca, NY.
- Seeks to define groups and the dominant species within.
- Not intended to find indicator species in pre-defined groups.
- Input community data, the analysis performs it's own clustering based on the assumption that indicator species are present
- Analysis is based on an initial CA
  - Inherits issues associated with CA (arch)
  - Assumption is there is one strong gradient.
  - Samples are divided along CA axes. As with DCA, cut points are arbitrary and criteria vague.

### Two Way Indicator Species (TWINSpan)

- Method for making divisions is complex, not well described in literature describing the technique.
  - At each iteration
    - Division of CA ordination is made
    - Minimal subset of species necessary to reproduce the ordination = indicator species
    - The division is refined by adjusting the weighting of species/samples
- There is no R code for doing this, the author of the package that does regular indicator species analysis considers TWINSpan to be fundamentally "broken".
- TWINSpan program:
  - <http://www.canodraw.com/wintwins.htm>

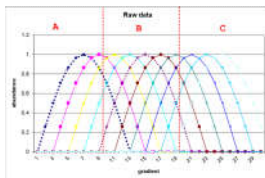


### Indicator Power of Species

- Question – "is species X an indicator of species Y", or "do species X and Y co-occur more than expected at random".
- Applications – facilitation, competitive interactions etc.
- First approach
  - Categorize samples based on the presence of species X, then perform traditional indicator species analysis
  - Abundances of species X not included, values would represent the relative frequency and abundance of species Y where X occurred.
  - See also [signassoc](#) function in indicpecies package.

### Indicator Power of Species

- Say we are interested in co-occurrence with species 3 (yellow)



```
classification<-community$sp3>0
```

```
ind_species<-multipatt(community,classification,duleg=TRUE)
```

```
ind_species
```

\$sign	e.FALSE	e.TRUE	index	stat	p.value
sp1	0	1	2	0.7076616	0.020
sp2	0	1	2	0.8753357	0.005
sp3	0	1	2	1.0000000	0.005
sp4	0	1	2	0.8753357	0.005
sp5	0	1	2	0.7076616	0.010
sp6	0	1	2	0.5030200	0.195
sp7	1	0	1	0.5651889	0.310
sp8	1	0	1	0.7247320	0.020
sp9	1	0	1	0.7817360	0.015
sp10	1	0	1	0.7817360	0.010

### Indicator Power of Species

- Function `indpower` in `vegan` package
  - Follows method of Halme et al (2009) and calculates pairwise indicator power.

```
> indpower(community)
Target Species
Indicator Species
t.sp1 t.sp2 t.sp3 t.sp4 t.sp5 t.sp6 t.sp7 t.sp8 t.sp9 t.sp10
i.sp1 1.0000000 0.8528029 0.7035265 0.5504819 0.3892495 0.2010076 0.0000000 0.0000000 0.0000000 0.0000000
i.sp2 0.8528029 1.0000000 0.8528029 0.7035265 0.5504819 0.3892495 0.2010076 0.0000000 0.0000000 0.0000000
i.sp3 0.7035265 0.8528029 1.0000000 0.8528029 0.7035265 0.5504819 0.3892495 0.2010076 0.0000000 0.0000000
i.sp4 0.5504819 0.7035265 0.8528029 1.0000000 0.8528029 0.7035265 0.5504819 0.3892495 0.2010076 0.0000000
i.sp5 0.3892495 0.5504819 0.7035265 0.8528029 1.0000000 0.8528029 0.7035265 0.5504819 0.3892495 0.2010076
i.sp6 0.2010076 0.3892495 0.5504819 0.7035265 0.8528029 1.0000000 0.8528029 0.7035265 0.5504819 0.3892495
i.sp7 0.0000000 0.2010076 0.3892495 0.5504819 0.7035265 0.8528029 1.0000000 0.8528029 0.7035265 0.5504819
i.sp8 0.0000000 0.0000000 0.2010076 0.3892495 0.5504819 0.7035265 0.8528029 1.0000000 0.8528029 0.7035265
i.sp9 0.0000000 0.0000000 0.0000000 0.2010076 0.3892495 0.5504819 0.7035265 0.8528029 1.0000000 0.8528029
i.sp10 0.0000000 0.0000000 0.0000000 0.0000000 0.2010076 0.3892495 0.5504819 0.7035265 0.8528029 1.0000000
```

Halme, P., Mönkkönen, M., Kotiaho, J. S., Yliäimö, A.-L. 2009. Quantifying the indicator power of an indicator species. *Conservation Biology* 23: 1008–1016.

### Indicator Species

- Characteristic species in discrete groups of samples
  - Indicator species
  - TWINSPAN
- Species associations
  - Indicator power of species
- Identification of species most responsible for differences among groups of samples
  - SIMPER (similarity percentage)
    - Based on abundance, does not weigh occurrence frequency as indicator species analysis does.

### SIMPER (similarity Percentage)

- Attempt to assess each species contribution to Bray-Curtis similarity among groups. For a given sample, one species contribution is:

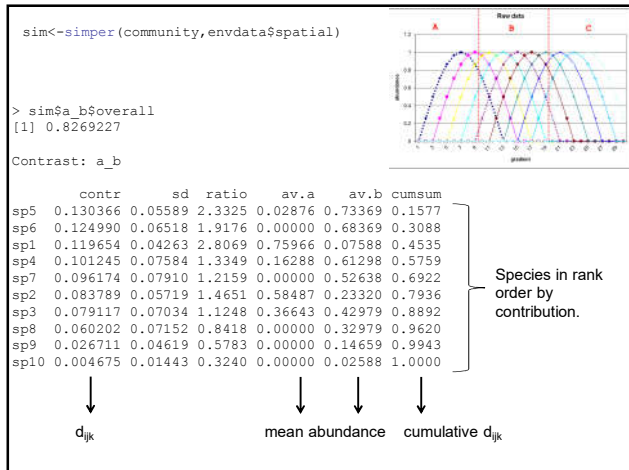
$$d_{ijk} = \frac{|(x_{ij} - x_{ik})|}{x_{ij} + x_{ik}}$$

- Where  $x$  is the abundance of species  $i$  in sample  $j$  and  $k$
- Average contribution for that species is the average across samples:

$$d_{jk} = \sum_{i=1}^S d_{ijk}$$

- Function `simper`(community, grouping)
  - limited to use with Bray Curtis index
  - Vegan package

Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18, 117–143



### Assignment

- Text: Section 4.10
- Reading:
  - Taylor, C.M. and M. E. Roberts. 2008. Using community-level analyses to identify dietary patterns for species in space and time. *Journal of Freshwater Ecology* 23:519-528.
  - Heino, J. et al. 2003. Defining macroinvertebrate assemblage types of headwater streams: implications for bioassessment and conservation. *Ecological Applications* 13:842-852.

### Assignment

- Tikus island coral reef dataset
  - library(mvabund)
  - data(tikus)
    - 10 transects sampled 6 years (81, 83-88). Coral bleaching event in 82-83.
    - Abundance of 75 coral species
- Categorize years as pre (81) and post (all other) years
  - What species are indicators for pre vs. post?
- Square-root transform data
  - Produce nmds plot (k=2, brak-Curtis distance)
  - Use simper to identify species most responsible for differences among years.
- Explain similarities/differences between approaches.

Warwick, R. M., & Clarke, K. R. (1990). A statistical analysis of coral community responses to the 1982-83 El Niño in the Thousand Islands, Indonesia. *Coral Reefs*, 8(4), 171-179.

