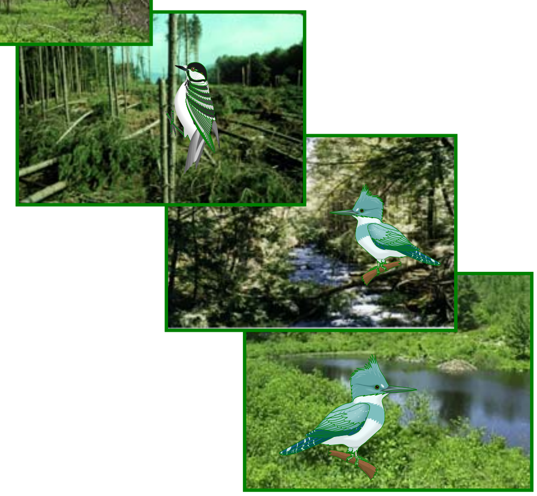
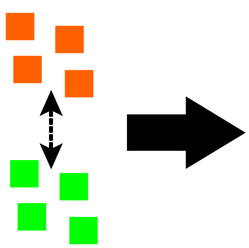
Discrimination Among Groups

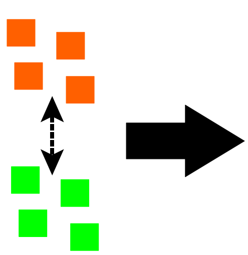


Id	Species	Canopy Cover	Snag Density	Canopy Height
1	A	80	1.2	35
2	A	75	0.5	32
3	A	72	2.8	28
.	.	.	.	.
31	B	35	3.3	15
32	B	75	4.1	25
60	B	15	5.0	3
.	.	.	.	.
61	C	5	2.1	5
62	C	8	3.4	2
90	C	25	0.6	15

Discrimination Among Groups



- Are groups significantly different? (How valid are the groups?)
  - Multivariate Analysis of Variance [(NP)MANOVA]
  - Multi-Response Permutation Procedures [MRPP]
  - Analysis of Group Similarities [ANOSIM]
  - Mantel's Test [MANTEL]

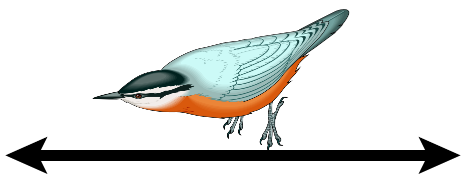


- How do groups differ? (Which variables best distinguish among the groups?)
  - Discriminant Analysis [DA]
  - Classification and Regression Trees [CART]
  - Logistic Regression [LR]
  - Indicator Species Analysis [ISA]

Tests of Among-Group Differences

2-group bird guilds example

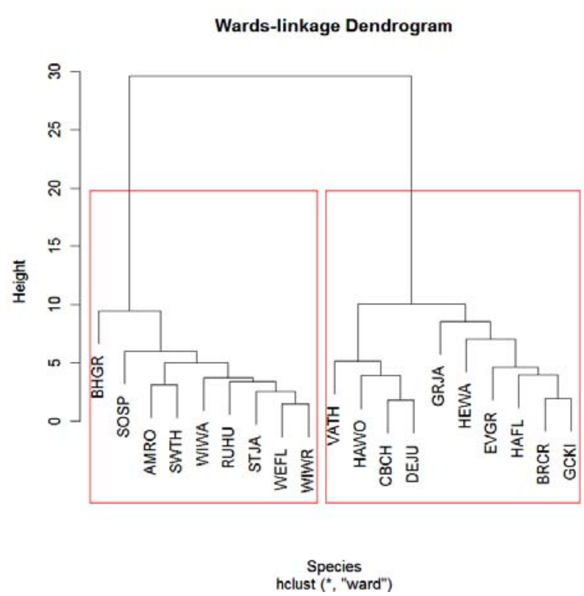
OBS	SPECIES	GTOTAL	LTOTAL	TTOTAL	MTOTAL	OTOTAL	SNAGM1	SNAGM23	SNAGM45	.	.	.	FHD
1	AMRO	15.31	31.42	64.28	20.71	47.14	0.00	0.28	0.14	.	.	.	1.45
2	BHGR	5.76	24.77	73.18	22.95	61.59	0.00	0.00	1.09	.	.	.	1.28
3	BRCR	4.78	64.13	30.85	12.03	63.60	0.44	0.44	2.08	.	.	.	1.18
4	CBCH	3.08	58.52	39.69	15.47	62.19	0.31	0.28	1.52	.	.	.	1.21
5	DEJU	13.90	60.78	36.50	13.81	62.89	0.23	0.31	1.23	.	.	.	1.23
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
19	WIWR	8.05	41.09	55.00	18.62	53.77	0.09	0.18	0.81	.	.	.	1.36



Tests of Among-Group Differences

2-group bird guilds example

- Two-group clustering solution resulting from agglomerative hierarchical clustering (hclust) using Euclidean distance on standardized variables and Ward's linkage method.



Tests of Among-Group Differences

Permutational Multivariate Analysis of Variance (NP-MANOVA)

- Nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities based on the analysis and partitioning sums of square distances (Anderson 2001).

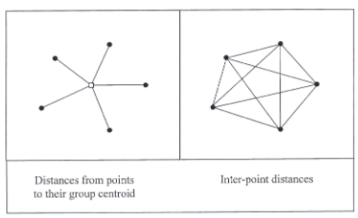


Fig. 2. The sum of squared distances from individual points to their centroid is equal to the sum of squared inter-point distances divided by the number of points.

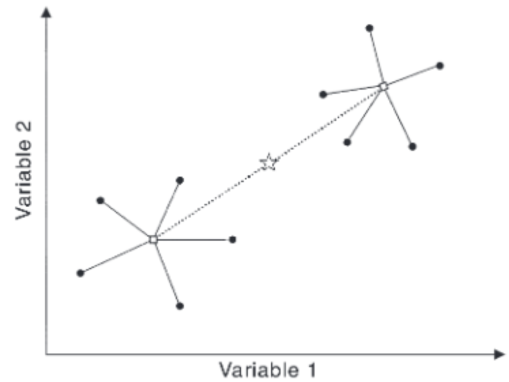


Fig. 1. A geometric representation of MANOVA for two groups in two dimensions where the groups differ in location. The within-group sum of squares is the sum of squared distances from individual replicates to their group centroid. The among-group sum of squares is the sum of squared distances from group centroids to the overall centroid. (——) Distances from points to group centroids; (·······) distances from group centroids to overall centroid; (☆), overall centroid; (□), group centroid; (●), individual observation.

Tests of Among-Group Differences

Permutational (nonparametric) MANOVA

- Calculate *distance* matrix (any distance metric can be used)
- Calculate average distance among all entities (SS_T)
- Calculate *average distance* among entities within groups (SS_W)
- Calculate average distance among groups ($SS_A = SS_T - SS_W$)
- Calculate *F-ratio*

$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2$$

$$SS_W = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}$$

$$F = \frac{SS_A/(a-1)}{SS_W/(N-a)}$$

N = total number of items
a = number of groups
n = number of items per group
d_{ij} = distance between ith and jth entity
ε_{ij} = 1 if in same group; 0 otherwise

Tests of Among-Group Differences

Permutational (nonparametric) MANOVA

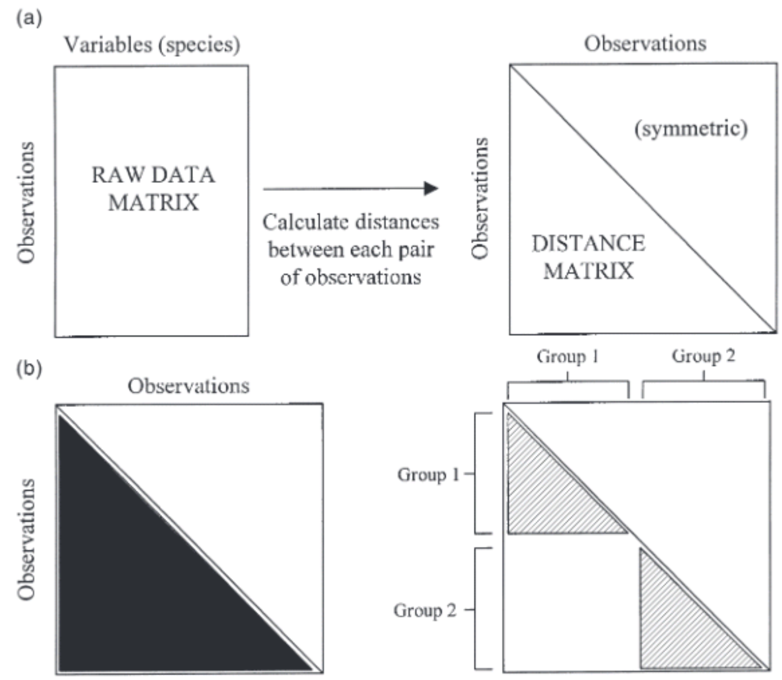
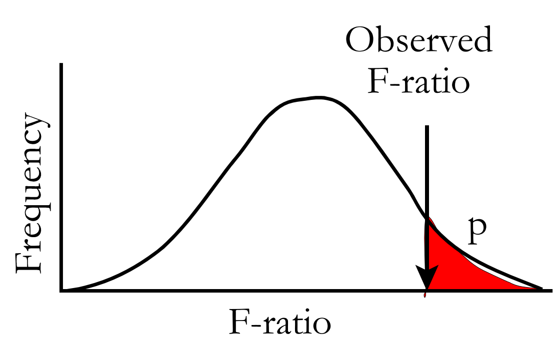


Fig. 3. Schematic diagram for the calculation of (a) a distance matrix from a raw data matrix and (b) a non-parametric MANOVA statistic for a one-way design (two groups) directly from the distance matrix. SS_T, sum of squared distances in the half matrix (■) divided by N (total number of observations); SS_W, sum of squared distances within groups (▨) divided by n (number of observations per group). $SS_A = SS_T - SS_W$ and $F = [SS_A/(a-1)]/[SS_W/(N-a)]$, where a = the number of groups.

Tests of Among-Group Differences

Permutational (nonparametric) MANOVA

- Determine the *probability* of an F-ratio this large or larger through Monte carlo permutations.
  - Permutations involve randomly assigning sample observations to groups (within strata if nested design).
  - The significance test is simply the fraction of permuted F-ratios that are greater than the observed F-ratio.



Tests of Among-Group Differences

Permutational (nonparametric) MANOVA

2-group bird guilds example

```
Call:
adonis(formula = y.std ~ grp, data = y.grp, permutations = 1000,
method = "euclidean")

Terms added sequentially (first to last)

          Df SumsOfSqs MeanSqs F.Model      R2   Pr(>F)
grp        1     171.6   171.6      13 0.43333 0.000999 ***
Residuals 17     224.4    13.2         0.56667
Total     18     396.0                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Conclude that the differences between the two bird species clusters is statistically significant ***and*** that 43% of the “variance” is accounted for by group differences

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

- Nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities based on permutation test of among- and within-group dissimilarities (Mielke 1984, 1991).

- Calculate *distance* matrix (Euclidean distance generally recommended, although proportional city-block measures often used with community data).
- Calculate *average distance* in each group $= \bar{d}_i$
- Calculate *delta* (the weighted mean within-group distance) for *g* groups.

$$delta = \delta = \sum_{i=1}^{g} C_i \bar{d}_i$$

$$C_i = {}^{n_i}\!/_{N}$$

n_i = number of items in group i
N = total number of items

**other options for calculating C_i exit*

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

2-group bird guilds example

Group		AMRO	BHGR	RUHU	SOSP	STJA	SWTH	WEFL	WIWA	WIWR	BRCR	CBCH	DEJU	EVGR	GCKI	GRJA	HAFL	HAWO	HEWA	VATH
1	AMRO	0																		
1	BHGR	6.83	0																	
1	RUHU	3.77	7.46	0																
1	SOSP	5.60	8.97	3.41	0															
1	STJA	4.88	7.26	3.14	4.56	0														
1	SWTH	3.11	4.70	3.34	5.57	3.54	0													
1	WEFL	3.61	6.57	3.27	4.85	2.58	2.51	0												
1	WIWA	5.13	6.68	3.49	5.42	3.46	3.45	3.44	0											
1	WIWR	3.71	6.78	2.40	4.25	2.02	2.49	1.50	2.56	0										
2	BRCR	9.15	9.29	7.44	7.65	5.98	7.52	6.66	5.18	6.18	0									
2	CBCH	7.04	7.68	5.14	5.66	3.84	5.28	4.50	3.14	3.85	2.70	0								
2	DEJU	6.98	8.20	4.87	5.35	4.11	5.61	4.94	3.49	4.10	3.11	1.82	0							
2	EVGR	10.73	10.57	8.53	8.57	7.70	9.14	8.66	6.43	7.95	4.19	4.82	5.24	0						
2	GCKI	9.62	9.32	7.58	7.51	6.37	7.92	7.17	5.74	6.68	1.97	3.13	3.40	4.08	0					
2	GRJA	10.22	9.71	9.68	9.60	8.50	8.93	8.25	7.85	8.23	5.30	5.88	6.20	8.06	5.53	0				
2	HAFL	11.21	10.18	9.15	9.32	8.07	9.44	9.00	6.97	8.48	3.98	5.10	5.46	3.98	2.96	7.29	0			
2	HAWO	7.49	7.69	5.59	5.83	4.54	5.70	5.38	4.49	4.67	4.19	3.29	3.49	6.30	4.30	7.36	6.02	0		
2	HEWA	12.42	11.75	10.53	10.09	9.70	11.17	10.70	9.24	10.22	6.41	7.47	7.32	6.86	5.17	8.25	4.56	8.04	0	
2	VATH	7.08	8.41	5.83	5.93	4.19	5.97	4.63	5.27	4.68	4.78	4.28	4.51	7.12	5.10	7.44	7.07	4.39	7.89	0

$\overline{d}_i = 4.34 \quad c_1 = 36/81 = .444$

$\overline{d}_2 = 5.24 \quad c_2 = 45/81 = .556$

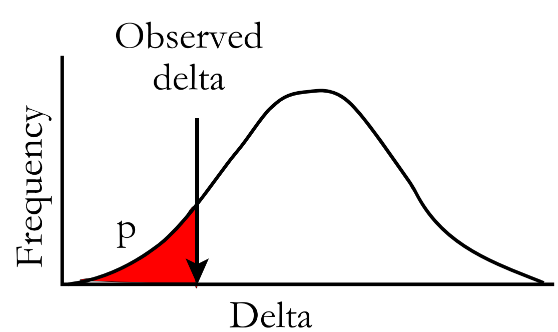
$\delta = 4.84$

Note: only with-group dissimilarities are used.

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

- Determine the *probability* of a delta this small or smaller through Monte carlo permutations.
  - Permutations involve randomly assigning sample observations to groups.
  - The significance test is simply the fraction of permuted deltas that are less than the observed delta, with a small sample correction.



Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

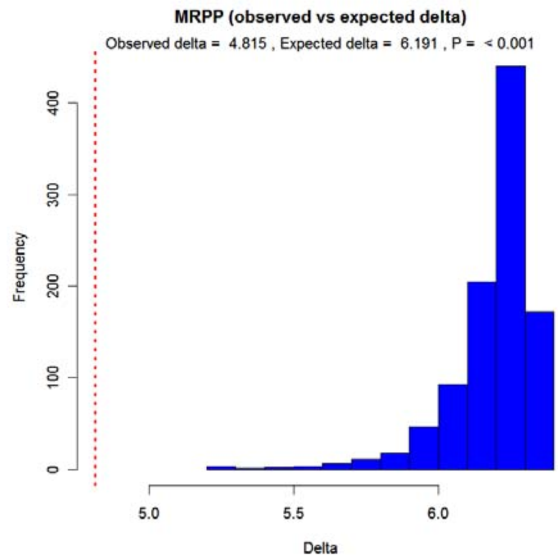
2-group bird guilds example

```
Call:
mrpp(dat = y.eucl, grouping = grp)

Dissimilarity index: euclidean
Weights for groups:  n

Chance corrected within-group agreement A: 0.2225
Based on observed delta 4.815 and expected delta 6.192

Significance of delta: < 0.001
Based on  1000  permutations
```



- Conclude that two clusters (bird species with similar niches) differ *significantly* in terms of the measured habitat variables.

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

- Determine the *effect size* independent of sample size (chance-corrected within-group agreement, *A*).

$$A = 1 - \frac{\delta}{\mu_\delta} = 1 - \frac{\text{observed } \delta}{\text{expected } \delta}$$

  - The statistic *A* is given as a descriptor of within-group homogeneity compared to the random expectation

    A = 1 when all items are identical within groups
    A = 0 when within-group heterogeneity equals expectation by chance
    A < .1 common in ecology
    A > .3 is fairly high in ecology (*but see simulation results later*)

Note: statistical significance (small p-value) may result even when the "effect size" (A) is small, if the sample size is large.

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

2-group bird guilds example

```
Call:
mrpp(dat = y.eucl, grouping = grp)

Dissimilarity index: euclidean
Weights for groups:  n

Chance corrected within-group agreement A: 0.2225
Based on observed delta 4.815 and expected delta 6.192

Significance of delta: < 0.001
Based on  1000  permutations
```

- Conclude that the differences between the two bird species clusters is statistically significant ***and*** that the difference is moderately large and therefore probably ecologically significant as well.

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Simulation Study

n=20, p=1, e=1

```
      grp          y           1
 [1,]   0 0.78172641 0.758466580
 [2,]   0 0.72322922 0.897645896
 [3,]   0 0.22718943 0.012008126
 [4,]   0 0.60753563 0.448890944
 [5,]   0 0.03840249 0.225436849
 [6,]   0 0.98348911 0.006262736
 [7,]   0 0.39210403 0.397803540
 [8,]   0 0.03179824 0.316845145
 [9,]   0 0.63247471 0.410760431
[10,]   0 0.73152952 0.178422687
[11,]   1 1.10285749 0.318455045
[12,]   1 1.10485233 0.612280473
[13,]   1 1.72840012 0.631006766
[14,]   1 1.37381699 0.801464925
[15,]   1 1.89480304 0.686331478
[16,]   1 1.52831037 0.047549311
[17,]   1 1.30526785 0.656755662
[18,]   1 1.29030220 0.209625201
[19,]   1 1.36418721 0.099046791
[20,]   1 1.37266970 0.253065173
```

n=10, p=1, e=1

```
      grp         y          1
 [1,]   0 0.2460432 0.67330864
 [2,]   0 0.3473905 0.23258885
 [3,]   0 0.5006673 0.95751857
 [4,]   0 0.0881539 0.79033950
 [5,]   0 0.4879530 0.37304795
 [6,]   1 1.8694004 0.69309441
 [7,]   1 1.9764505 0.81360903
 [8,]   1 1.0345481 0.57213216
 [9,]   1 1.0238596 0.79868768
[10,]   1 1.3103284 0.06387041
```

n=10, p=1, e=4

```
      grp         y         1
 [1,]   0 0.6511912 0.8109652
 [2,]   0 0.4497334 0.4843661
 [3,]   0 0.7441406 0.7170146
 [4,]   0 0.2805700 0.3817845
 [5,]   0 0.7382498 0.2916249
 [6,]   1 4.2047604 0.2468846
 [7,]   1 4.4636355 0.9400324
 [8,]   1 4.3708406 0.4835018
 [9,]   1 4.3617041 0.1762655
[10,]   1 4.8397717 0.5176440
```

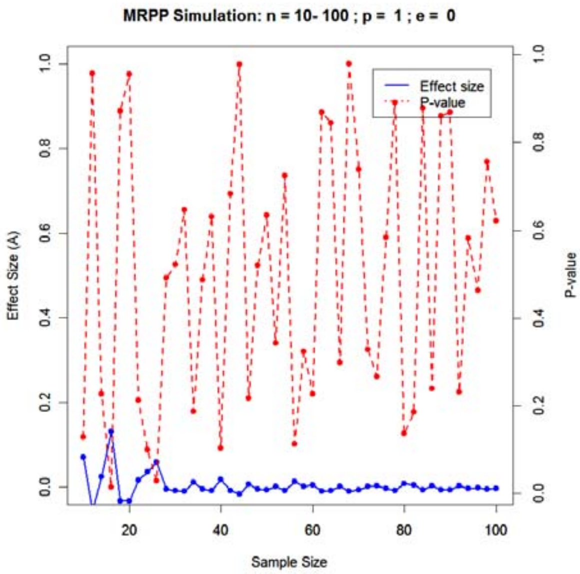
n=10, p=3, e=1

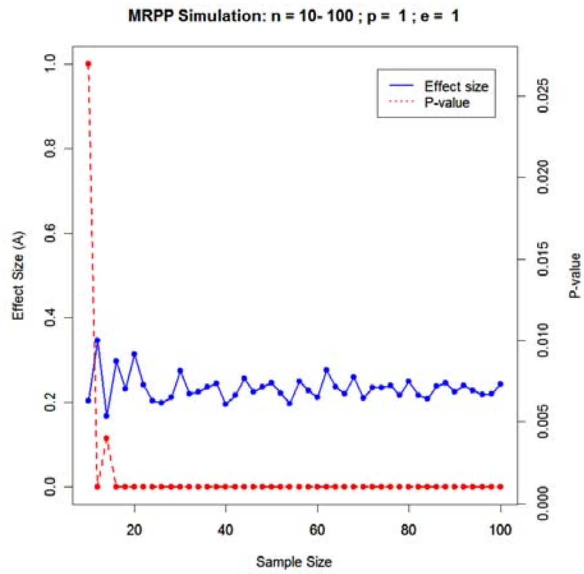
```
      grp         y         1         2          3
 [1,]   0 0.2384904 0.9270140 0.1637725 0.04744548
 [2,]   0 0.5240702 0.4438856 0.8387992 0.73323831
 [3,]   0 0.4846718 0.0880790 0.5378692 0.66965011
 [4,]   0 0.7693610 0.1904204 0.8995969 0.26179045
 [5,]   0 0.1217240 0.5295393 0.6956824 0.79223646
 [6,]   1 1.7356405 0.7805704 0.7213608 0.17149252
 [7,]   1 1.4519720 0.1926339 0.8744577 0.48857552
 [8,]   1 1.8991477 0.8895194 0.5275195 0.17056135
 [9,]   1 1.7226480 0.3448639 0.0082422 0.29317219
[10,]   1 1.1669151 0.3416431 0.9159500 0.63253065
```

Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

e=0 Sample size e=1

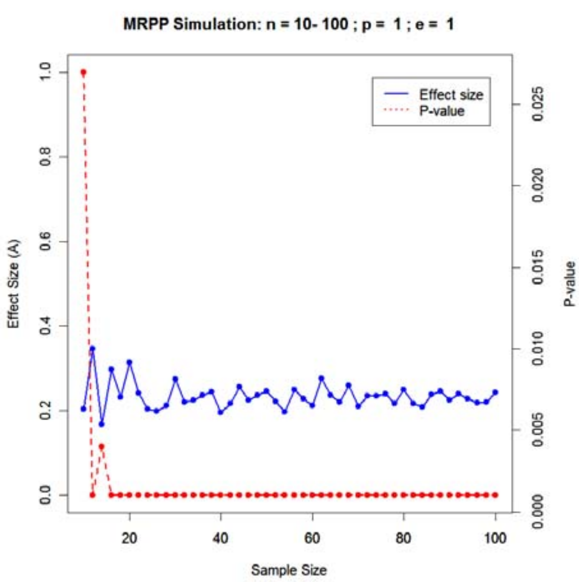


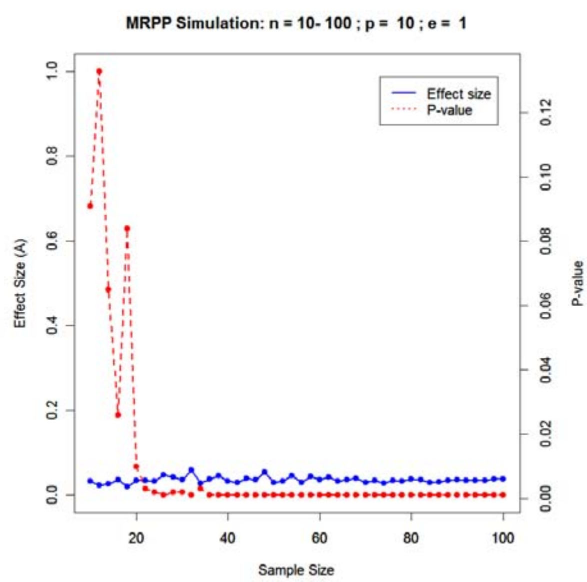


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

p=1 Sample size p=10

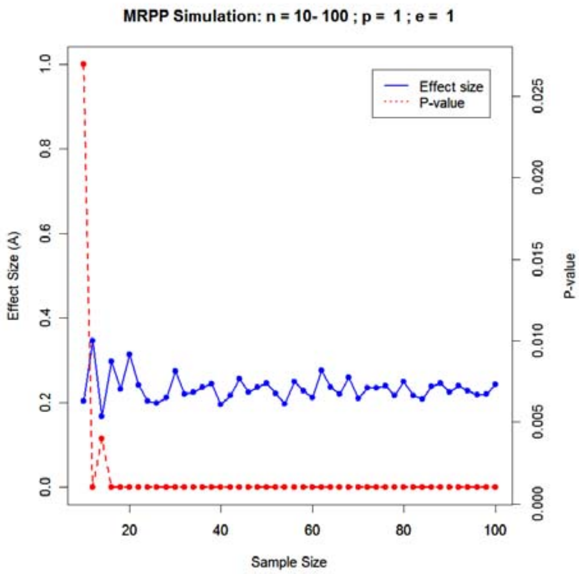


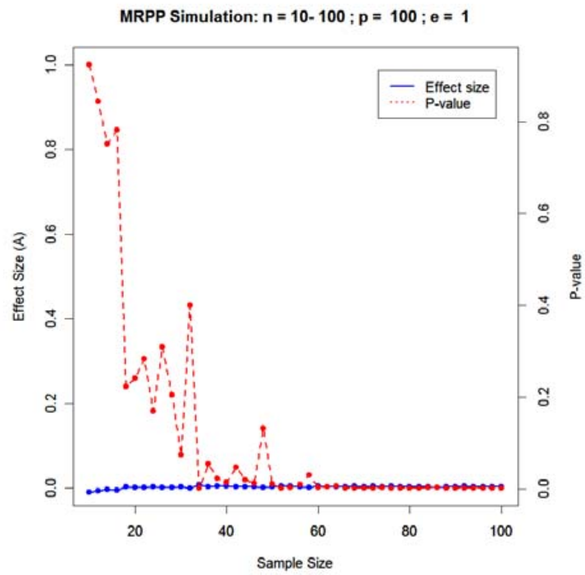


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

p=1 Sample size p=100

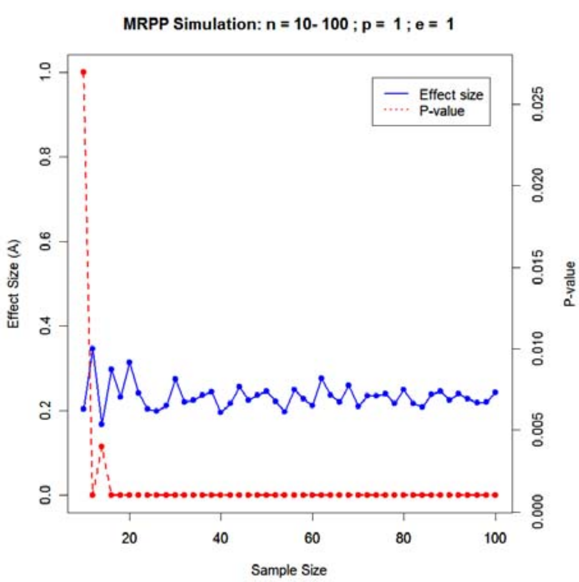


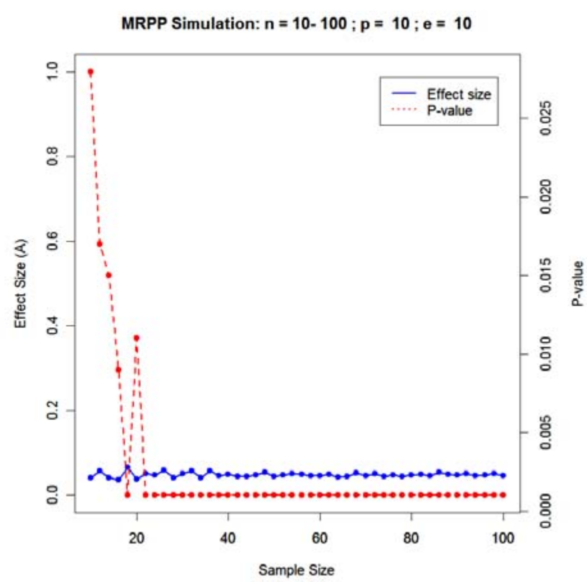


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

p=1, e=1 Sample size p=10, e=10

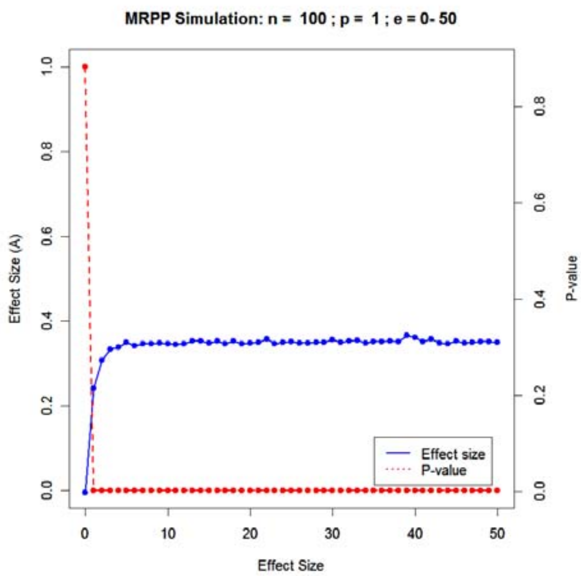


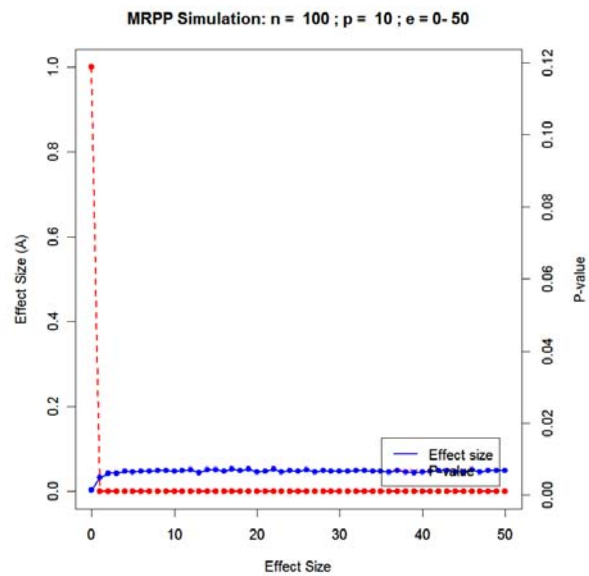


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

p=1 Effect size p=10

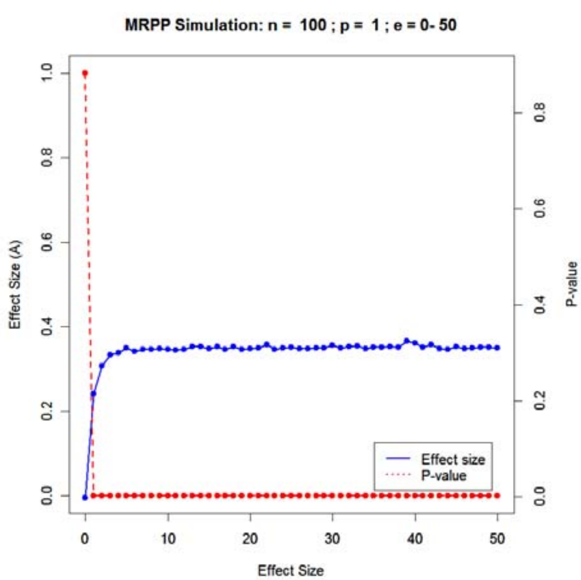


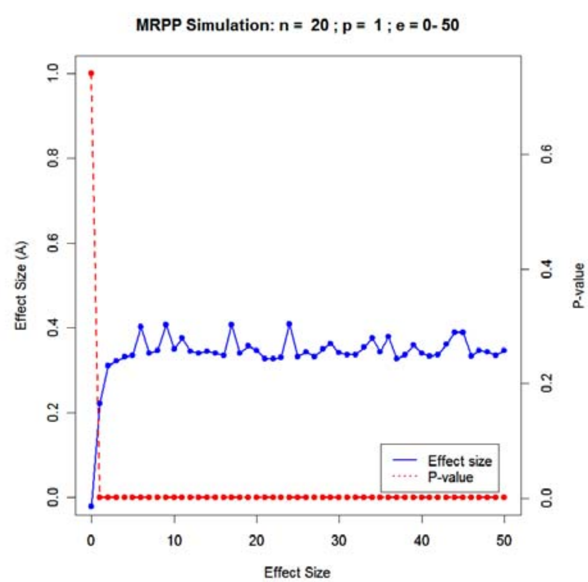


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

n=100 Effect size n=20

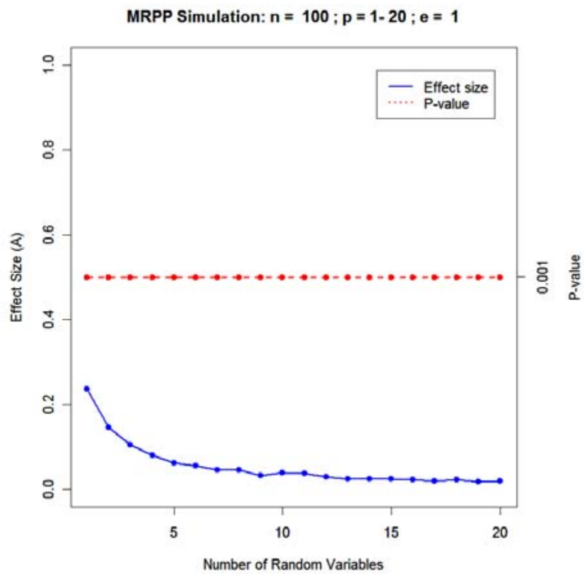


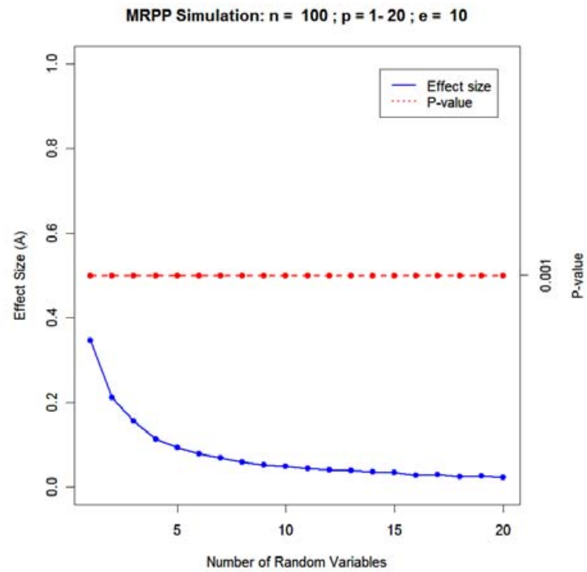


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

e=1 Noise Ratio e=10

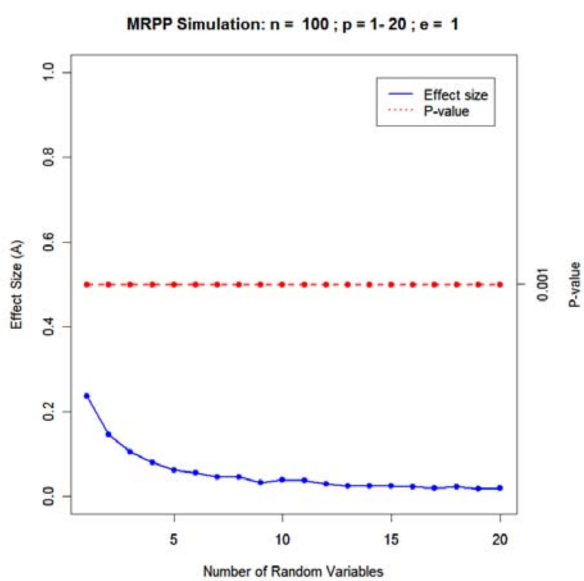


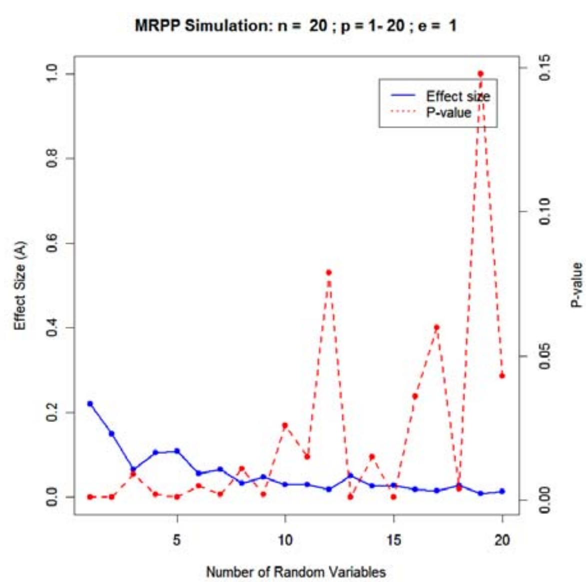


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

n=100 Noise Ratio n=20

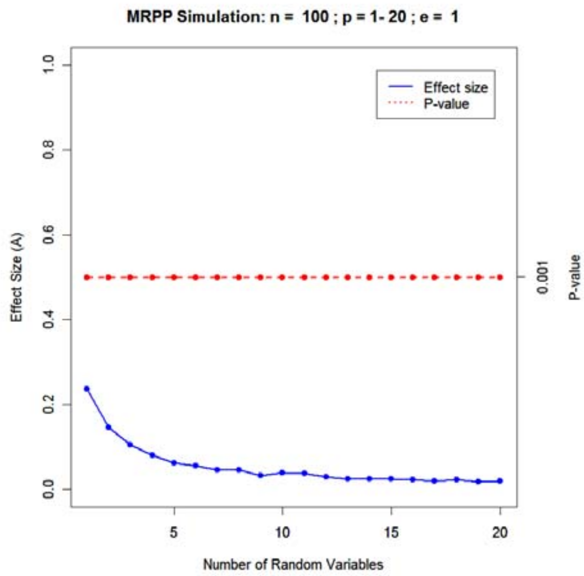


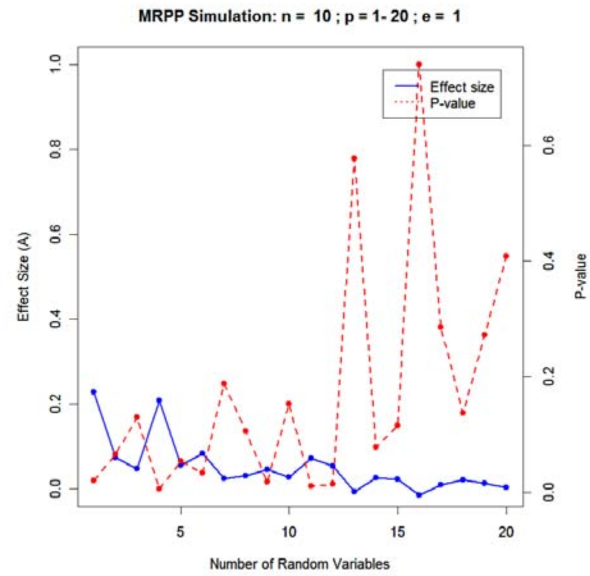


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

n=100 Noise Ratio n=10

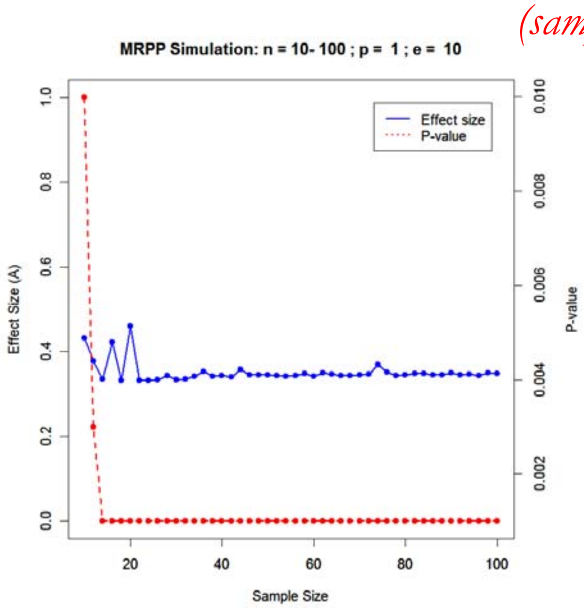


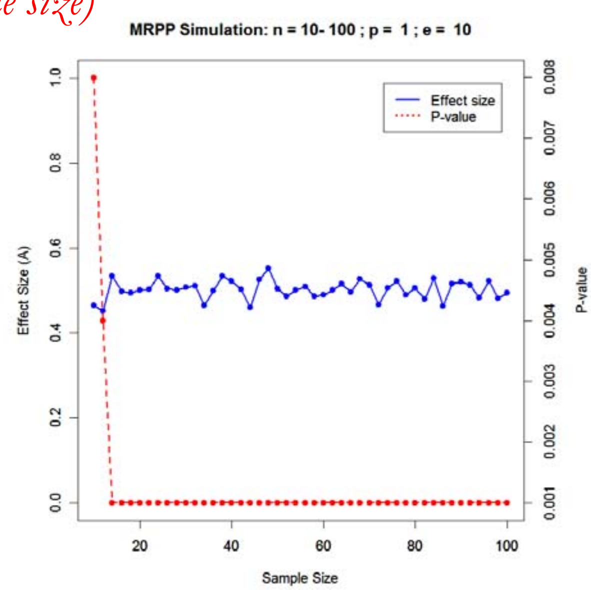


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Column standardized Data Standardization *(sample size)* Raw (same range)

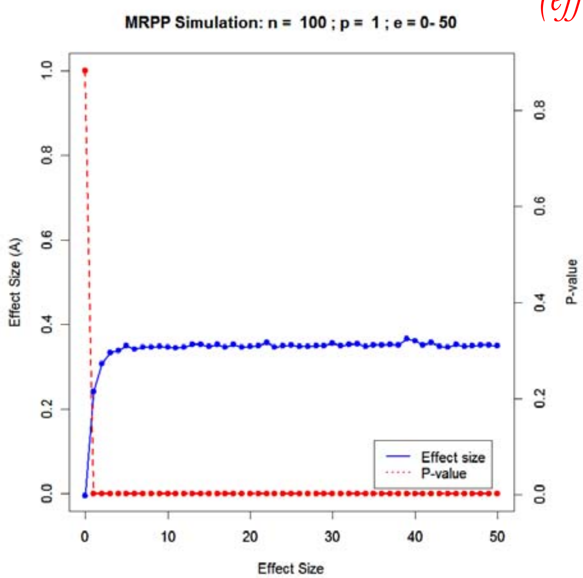


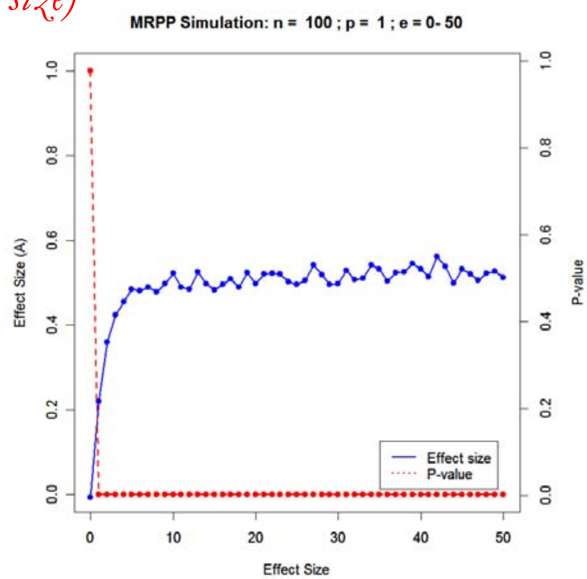


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Column standardized | Data Standardization *(effect size)* | Raw (same range)

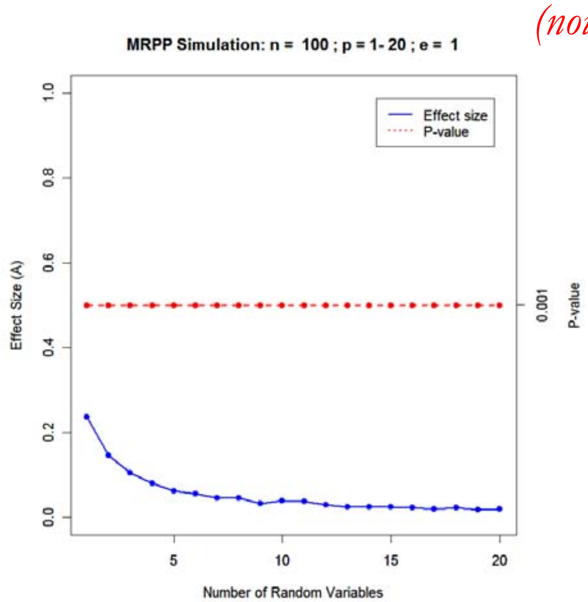


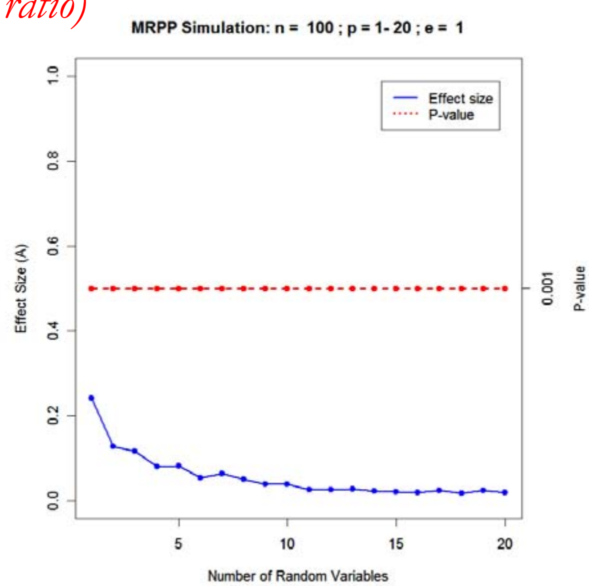


Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Column standardized | Data Standardization *(noise ratio)* | Raw (same range)





Tests of Among-Group Differences

Multi-Response Permutation Procedures (MRPP)

Simulation Results Summary

- Test statistic is unreliable for *n<20*, and for larger sample sizes (e.g., n=40) if the noise ratio is greater than ~10:1.
- *Significance* (p-value) of test statistic insensitive to degree of noise (nondiscriminating variables); i.e., a single effective discriminating variable will produce a significant test result for n>20, as long as noise ratio <~10:1.
- *Effect size* (A) sensitive to both magnitude of effect on at least one discriminating variable and the noise ratio (i.e., ratio of nondiscriminating variables to true discriminators); noise ratio relatively more important.
- *Data standardization* (col z-score) has moderate effect on effect size (A).

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

- Nonparametric procedure for testing the hypothesis of no difference between two or more groups of entities based on permutation test of among- and within-group similarities (Clark 1993).

- Calculate *dissimilarity* matrix.
- Calculate *rank* dissimilarities (smallest dissimilarity is given a rank of 1).
- Calculate mean among- and within-group rank dissimilarities.
- Calculate *test statistic* R (an index of relative within-group dissimilarity).

$$R = \frac{\bar{r}_A - \bar{r}_W}{M/2}$$

M = N(N-1)/2
= number of sample pairs

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

R is interpreted like a correlation coefficient and is a measure of 'effect size', like *A* in MRPP:

$R = 1$ when all pairs of samples within groups are more similar than to any pair of samples from different groups.

$R = 0$ expected value under the null model that among-and within-group dissimilarities are the same on average.

$R < 0$ numerically possible but ecologically unlikely.

$$R = \frac{\bar{r}_A - \bar{r}_W}{M/2}$$

M = N(N-1)/2
= number of sample pairs

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

2-group bird guilds example

Group		AMRO	BHGR	RUHU	SOSP	STJA	SWTH	WEFL	WIWA	WIWR	BRCR	CBCH	DEJU	EVGR	GCKI	GRJA	HAFL	HAWO	HEWA	VATH
1	AMRO																			
1	BHGR	102																		
1	RUHU	31	117																	
1	SOSP	79	145	21																
1	STJA	60	111	16	51															
1	SWTH	13	55	19	77	28														
1	WEFL	29	97	17	58	9	7													
1	WIWA	64	99	27	74	24	23	22												
1	WIWR	30	101	5	42	4	6	1	8											
2	BRCR	148	151	116	123	89	121	98	67	91										
2	CBCH	106	124	65	81	32	70	47	15	33	10									
2	DEJU	105	134	59	72	38	80	61	25	37	12	2								
2	EVGR	167	165	141	142	126	147	143	96	130	41	57	68							
2	GCKI	156	152	122	120	94	129	110	83	100	3	14	20	36						
2	GRJA	162	159	157	155	140	144	137	127	135	71	86	92	132	76					
2	HAFL	169	161	149	153	133	154	146	104	139	34	62	75	35	11	112				
2	HAWO	119	125	78	85	49	82	73	46	53	40	18	26	93	44	114	90			
2	HEWA	171	170	164	160	158	168	166	150	163	95	118	113	103	66	136	50	131		
2	VATH	108	138	84	87	39	88	52	69	54	56	43	48	109	63	115	107	45	128	

$\bar{r}_a = 111.50 \quad \bar{r}_w = 57.68$

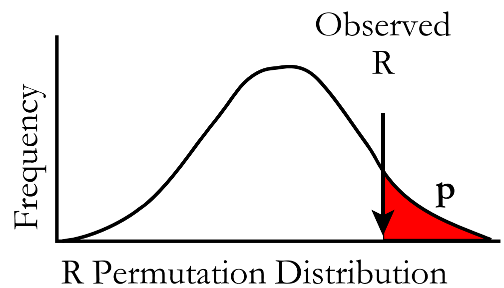
$R = 0.629$

Note: both within- and among-group dissimilarities are used.

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

- Determine the *probability* of an *R* this large or larger through Monte carlo permutations.
  - Permutations involve randomly assigning sample observations to groups.
  - The significance test is simply the fraction of permuted *R*'s that are greater than the observed *R*.



$$-1 \leq R \leq 1$$

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

2-group bird guilds example

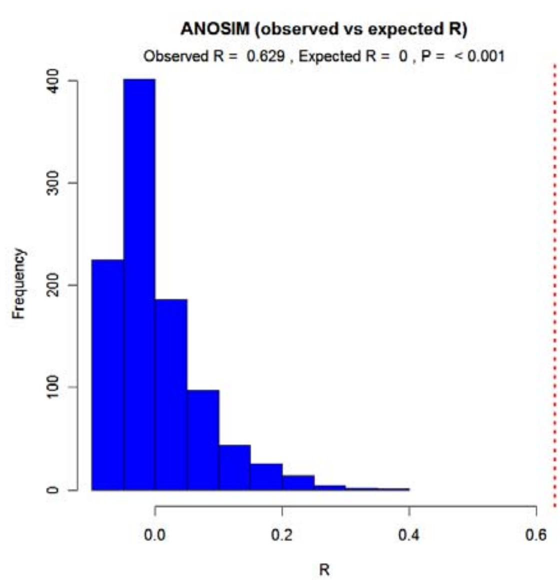
```
Call:
anosim(dis = y.eucl, grouping = grp)
Dissimilarity: euclidean

ANOSIM statistic R: 0.6294
      Significance: < 0.001

Based on  1000  permutations

Empirical upper confidence limits of R:
   90%    95%  97.5%    99%
0.0887 0.1397 0.1936 0.2573

Dissimilarity ranks between and within classes:
        0%   25%   50%    75% 100%  N
Between 15 80.25 121.5 148.75  171 90
1        1 16.75  29.5  74.75  145 36
2        2 36.00  63.0 103.00  136 45
```



- Conclude that two clusters differ *significantly* in terms of the measured habitat variables.

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Simulation Study

n=20, p=1, e=1

	grp	y	1
[1,]	0	0.78172641	0.758466580
[2,]	0	0.72322922	0.897645896
[3,]	0	0.22718943	0.012008126
[4,]	0	0.60753563	0.448890944
[5,]	0	0.03840249	0.225436849
[6,]	0	0.98348911	0.006262736
[7,]	0	0.39210403	0.397803540
[8,]	0	0.03179824	0.316845145
[9,]	0	0.63247471	0.410760431
[10,]	0	0.73152952	0.178422687
[11,]	1	1.10285749	0.318455045
[12,]	1	1.10485233	0.612280473
[13,]	1	1.72840012	0.631006766
[14,]	1	1.37381699	0.801464925
[15,]	1	1.89480304	0.686331478
[16,]	1	1.52831037	0.047549311
[17,]	1	1.30526785	0.656755662
[18,]	1	1.29030220	0.209625201
[19,]	1	1.36418721	0.099046791
[20,]	1	1.37266970	0.253065173

n=10, p=1, e=1

	grp	y	1
[1,]	0	0.2460432	0.67330864
[2,]	0	0.3473905	0.23258885
[3,]	0	0.5006673	0.95751857
[4,]	0	0.0881539	0.79033950
[5,]	0	0.4879530	0.37304795
[6,]	1	1.8694004	0.69309441
[7,]	1	1.9764505	0.81360903
[8,]	1	1.0345481	0.57213216
[9,]	1	1.0238596	0.79868768
[10,]	1	1.3103284	0.06387041

n=10, p=1, e=4

	grp	y	1
[1,]	0	0.6511912	0.8109652
[2,]	0	0.4497334	0.4843661
[3,]	0	0.7441406	0.7170146
[4,]	0	0.2805700	0.3817845
[5,]	0	0.7382498	0.2916249
[6,]	1	4.2047604	0.2468846
[7,]	1	4.4636355	0.9400324
[8,]	1	4.3708406	0.4835018
[9,]	1	4.3617041	0.1762655
[10,]	1	4.8397717	0.5176440

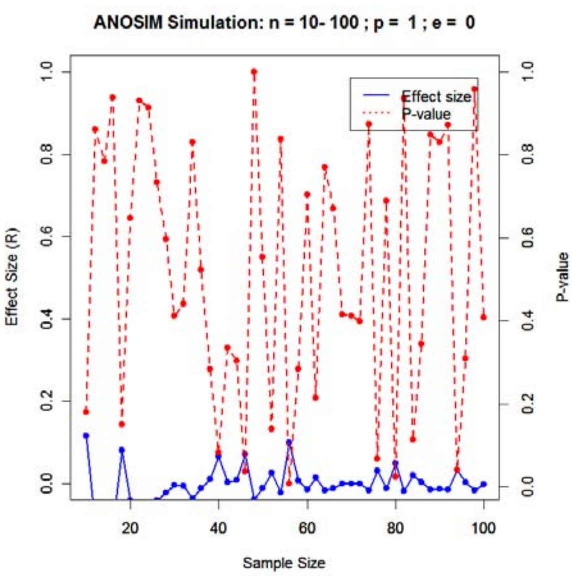
n=10, p=3, e=1

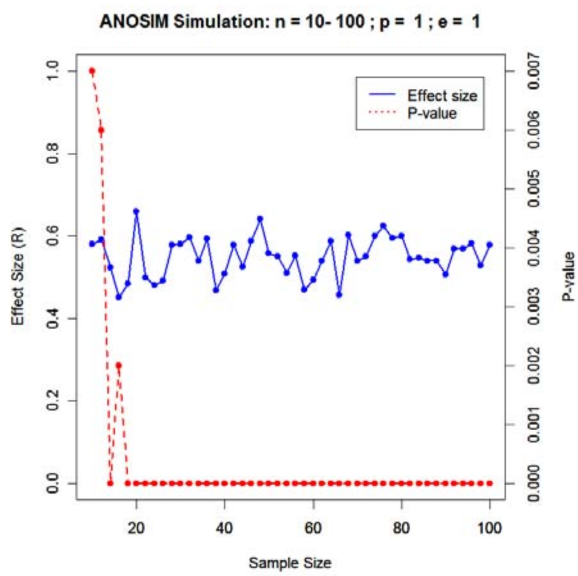
	grp	y	1	2	3
[1,]	0	0.2384904	0.9270140	0.1637725	0.04744548
[2,]	0	0.5240702	0.4438856	0.8387992	0.73323831
[3,]	0	0.4846718	0.0880790	0.5378692	0.66965011
[4,]	0	0.7693610	0.1904204	0.8995969	0.26179045
[5,]	0	0.1217240	0.5295393	0.6956824	0.79223646
[6,]	1	1.7356405	0.7805704	0.7213608	0.17149252
[7,]	1	1.4519720	0.1926339	0.8744577	0.48857552
[8,]	1	1.8991477	0.8895194	0.5275195	0.17056135
[9,]	1	1.7226480	0.3448639	0.0082422	0.29317219
[10,]	1	1.1669151	0.3416431	0.9159500	0.63253065

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

e=0 Sample size e=1

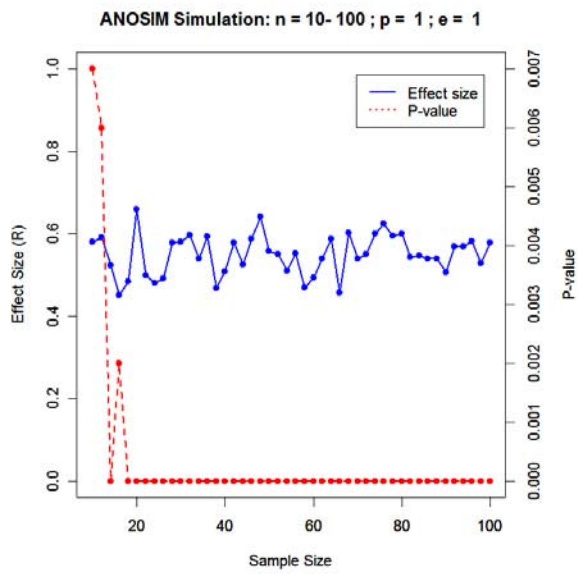


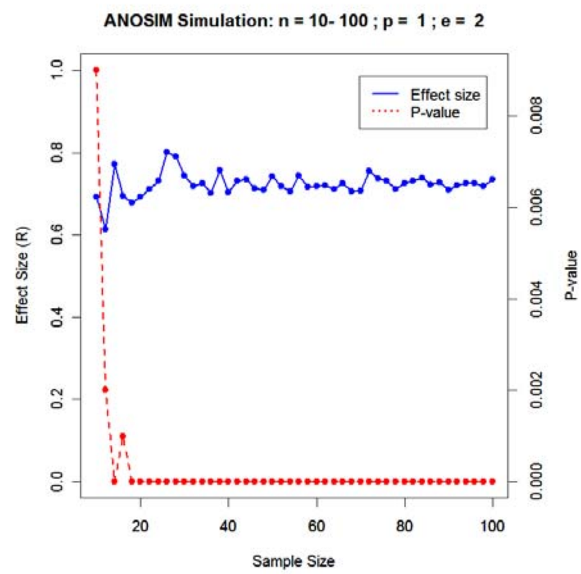


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

e=1 Sample size e=2

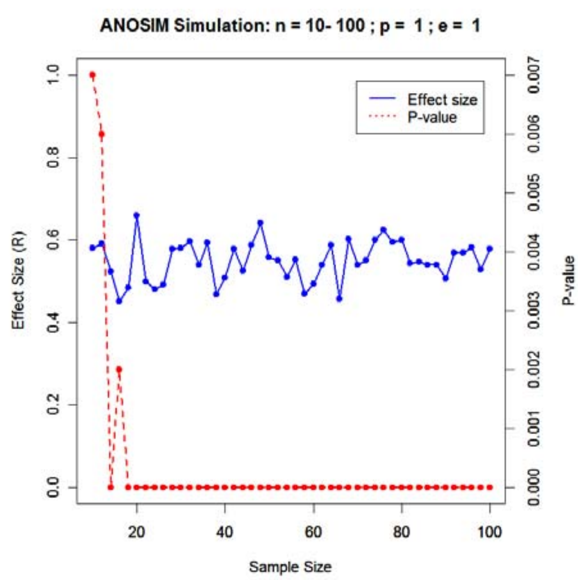


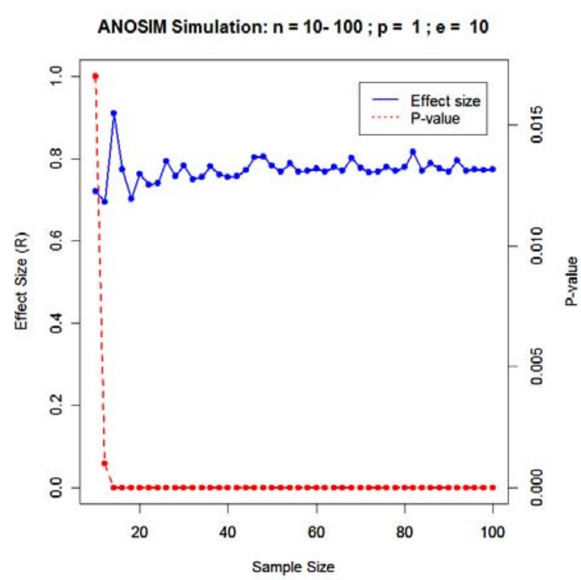


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

e=1 Sample size e=10

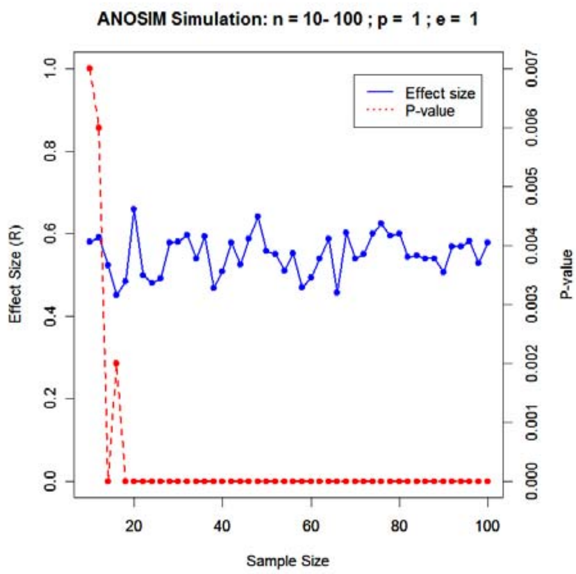


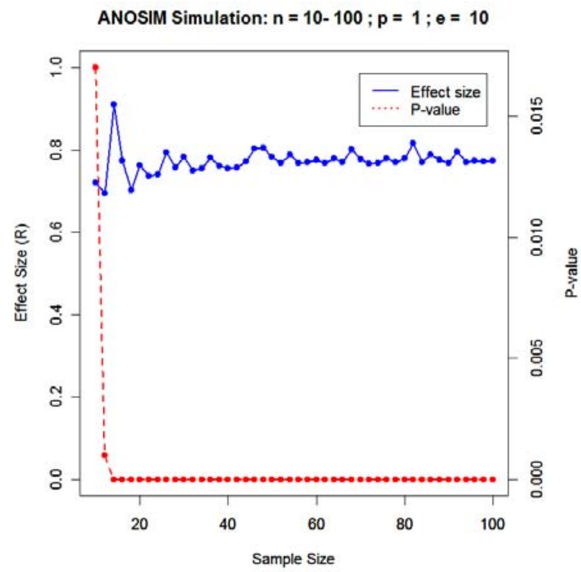


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

e=1 Sample size e=10

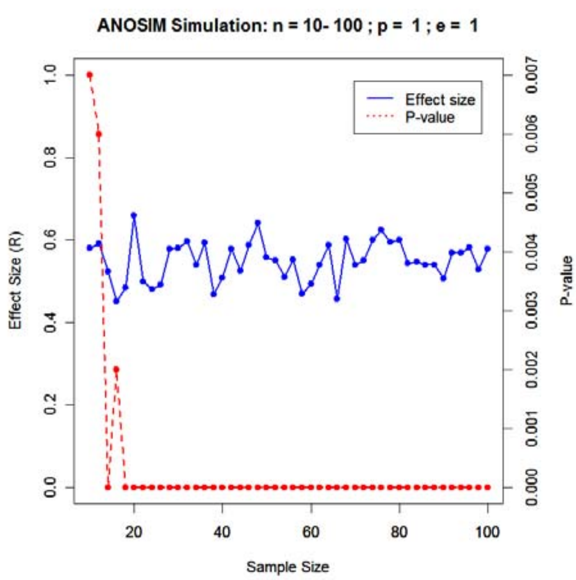


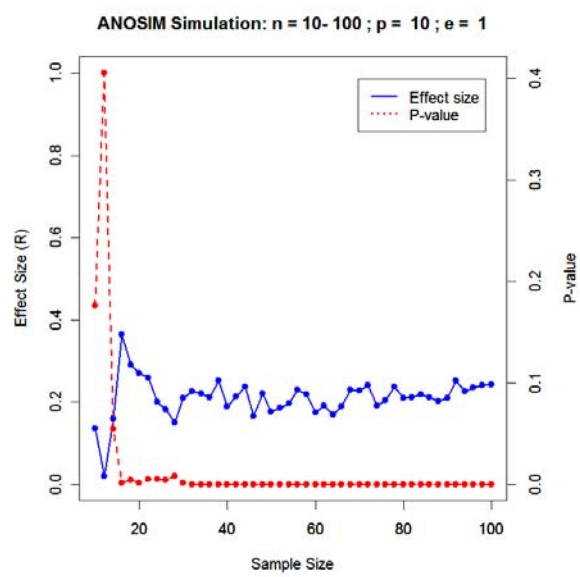


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

p=1 Sample size p=10

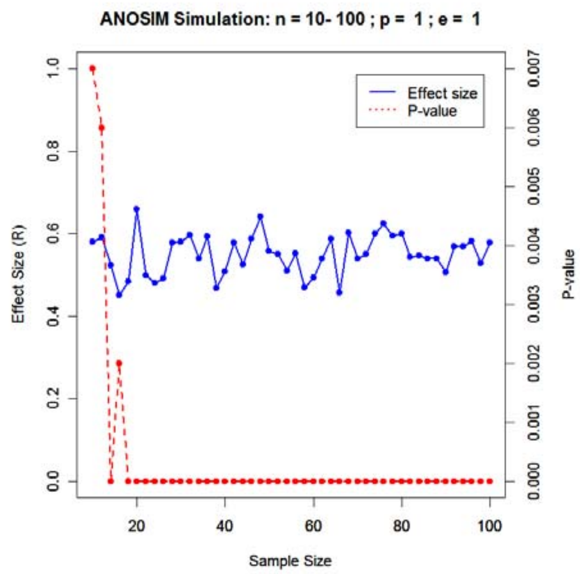


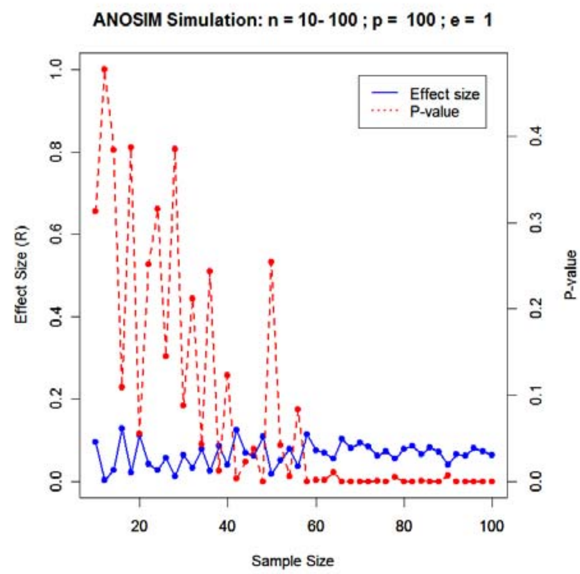


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

p=1 Sample size p=100



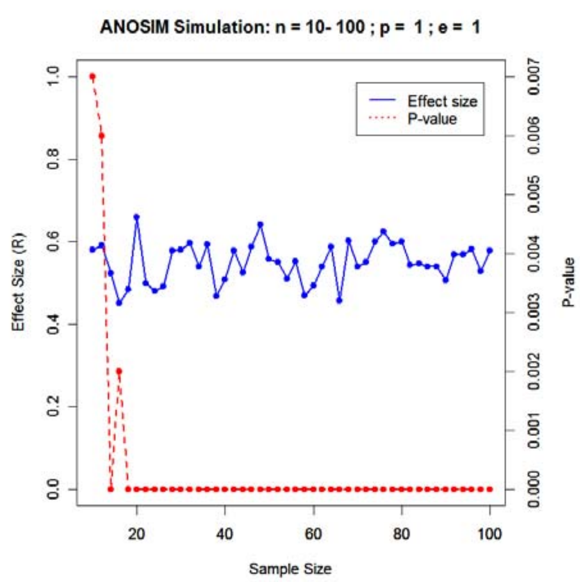


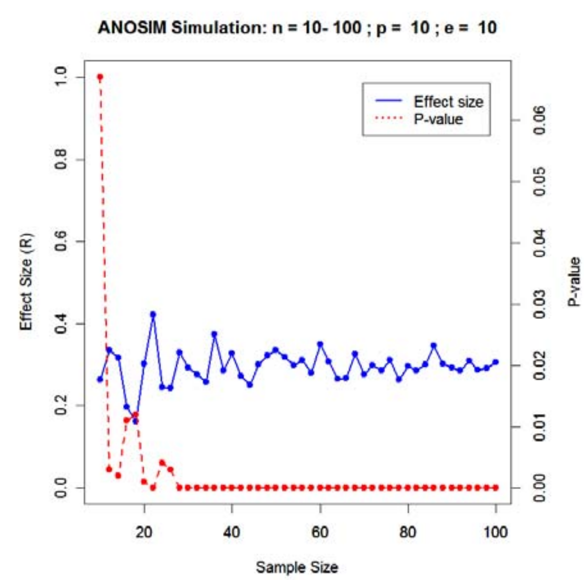
41

Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

p=1, e=1 Sample size p=10, e=10

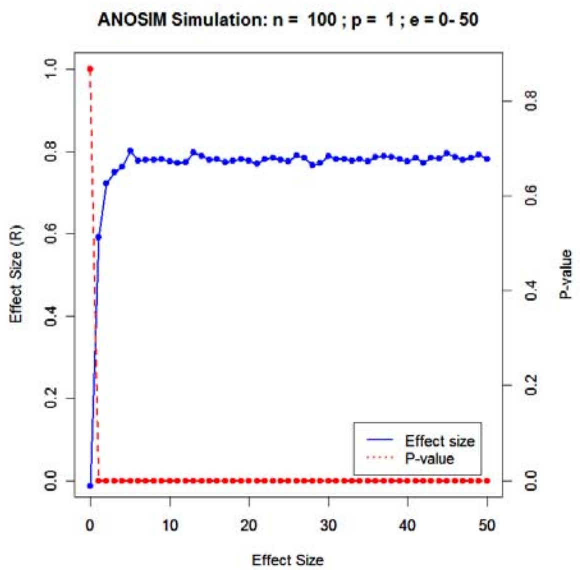


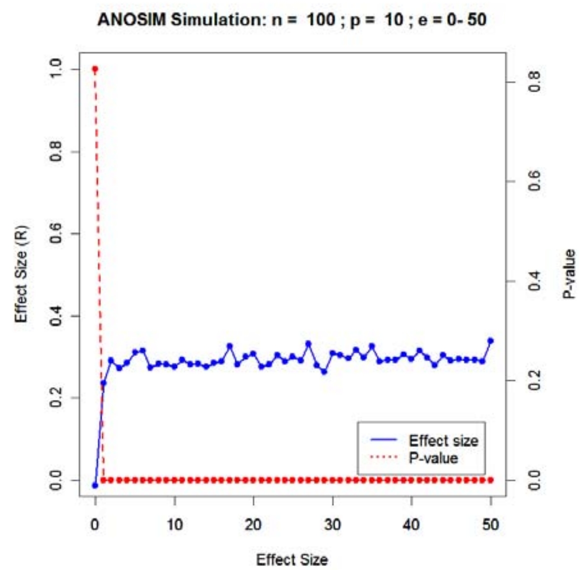


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

p=1 Effect size p=10

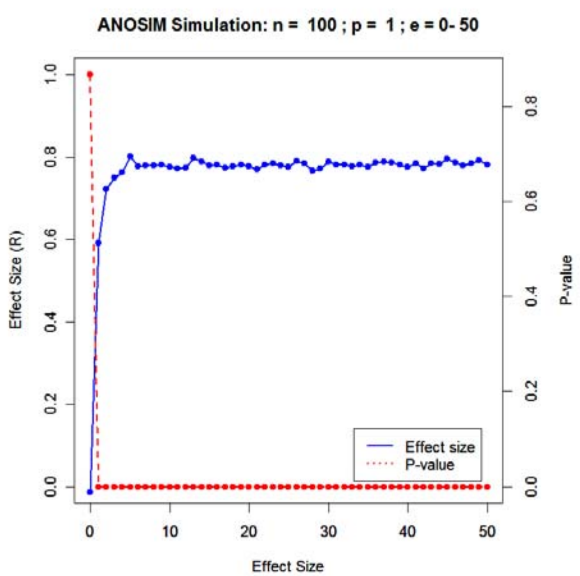


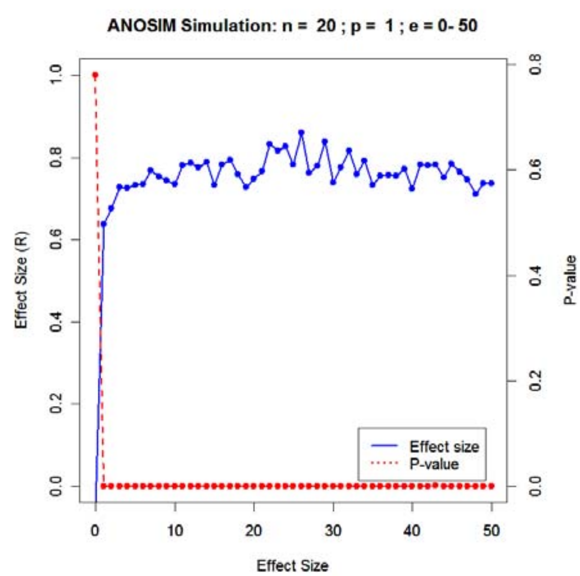


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

n=100 Effect size n=20

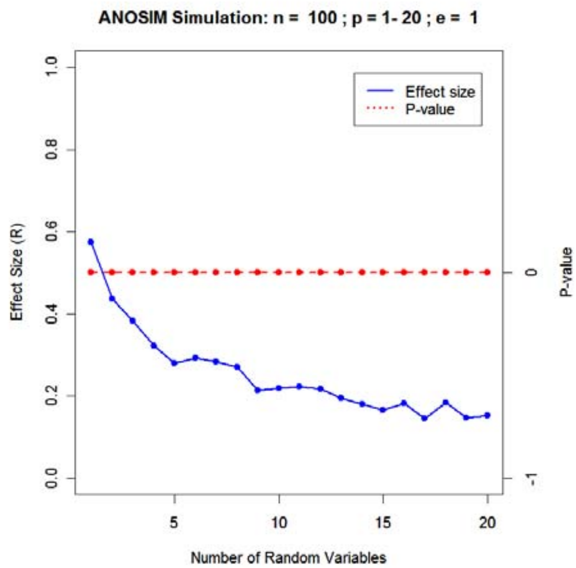


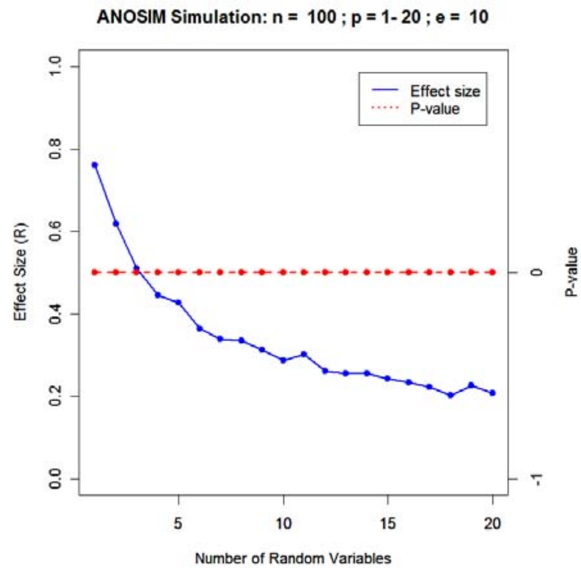


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

e=1 Noise Ratio e=10

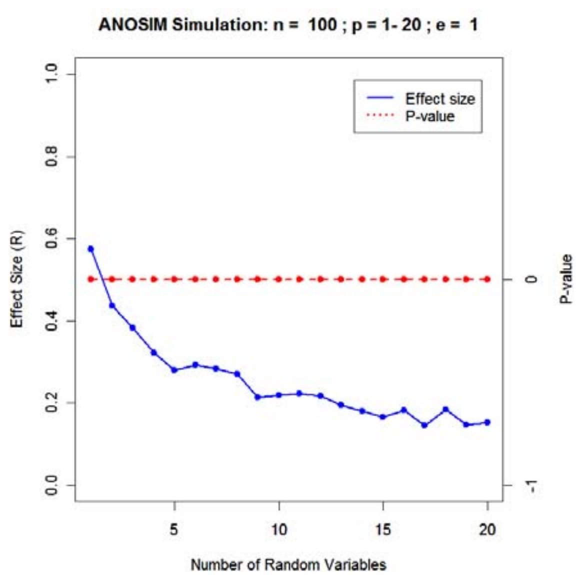


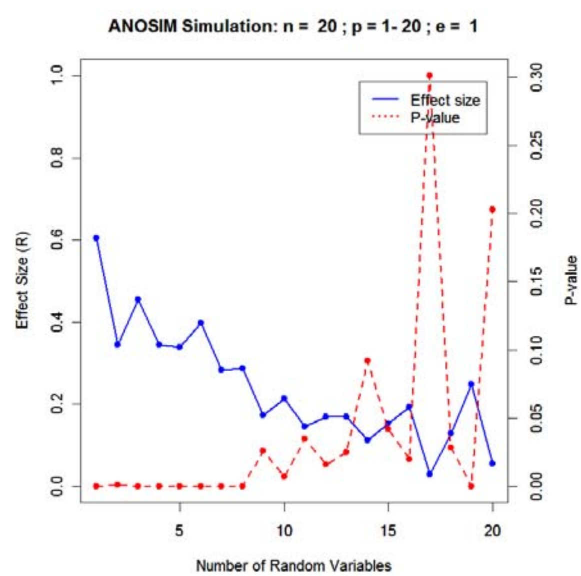


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

n=100 Noise Ratio n=20

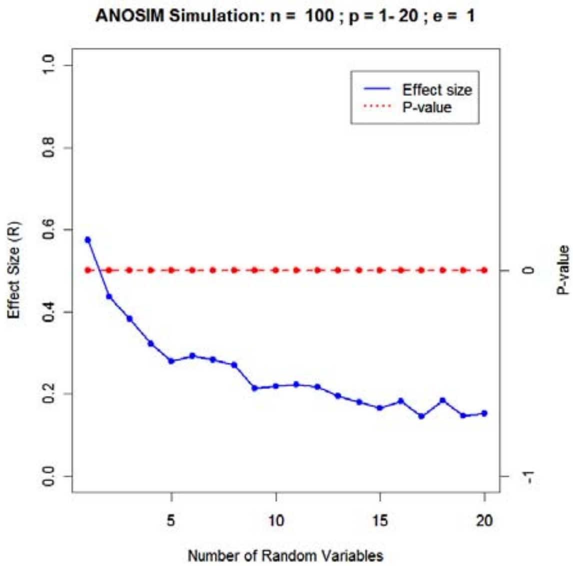


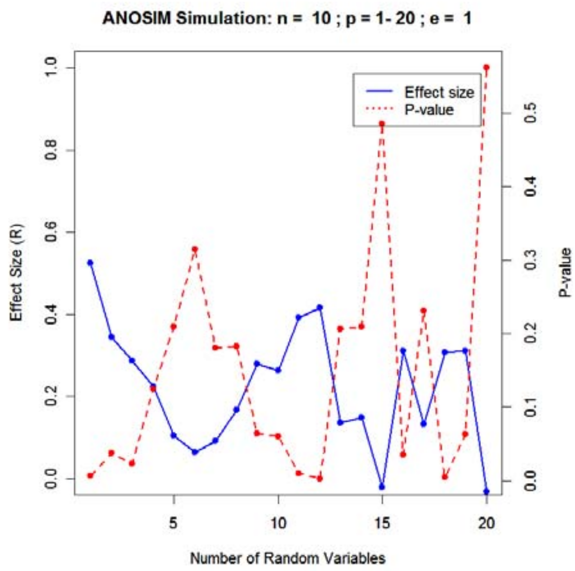


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

n=100 Noise Ratio n=10

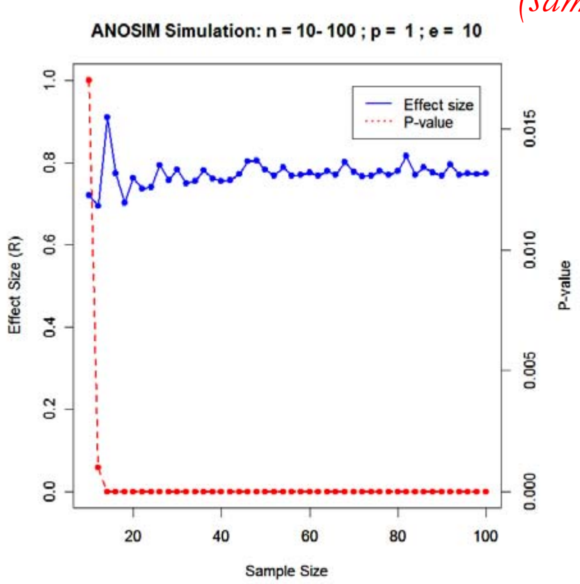


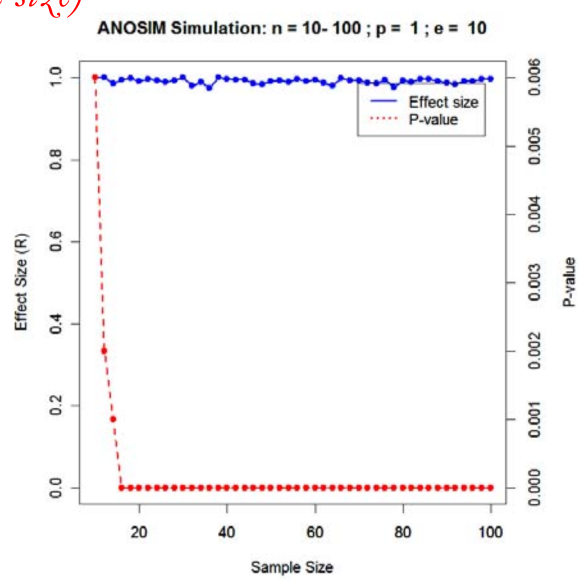


Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Column standardized Data Standardization *(sample size)* Raw (same range)





Tests of Among-Group Differences

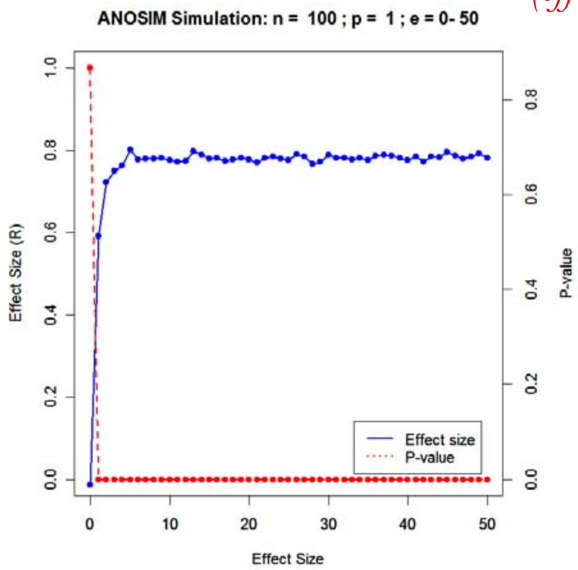
Analysis of Group Similarities (ANOSIM)

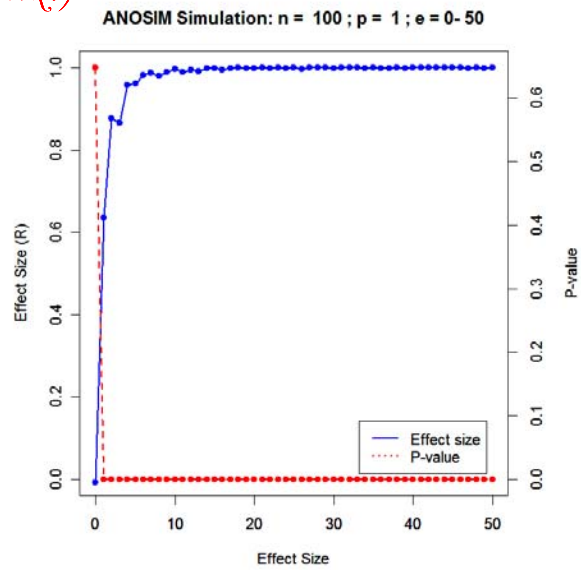
Column standardized

Data Standardization

(effect size)

Raw (same range)





Tests of Among-Group Differences

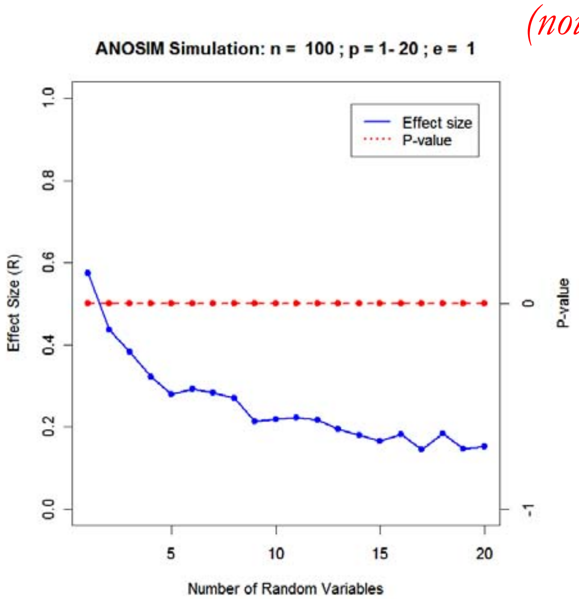
Analysis of Group Similarities (ANOSIM)

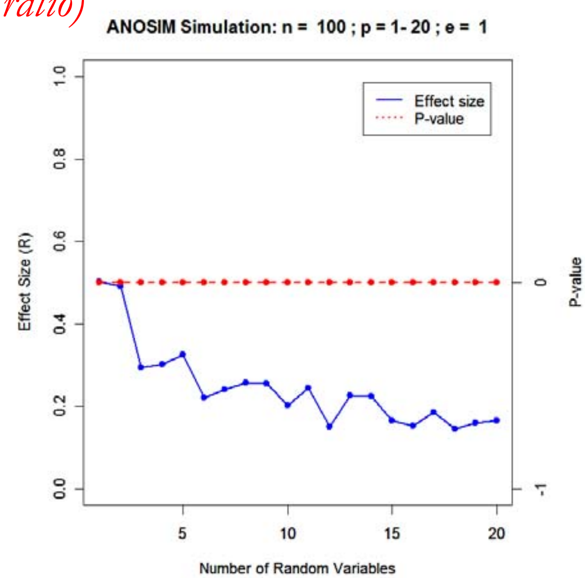
Column standardized

Data Standardization

(noise ratio)

Raw (same range)





Tests of Among-Group Differences

Analysis of Group Similarities (ANOSIM)

Simulation Results Summary

- Test statistic is unreliable for *n<20*, and for larger sample sizes (e.g., n=40) if the noise ratio is greater than ~10:1.
- *Significance* (p-value) of test statistic insensitive to degree of noise (nondiscriminating variables); i.e., a single effective discriminating variable will produce a significant test result for n>20, as long as noise ratio <~10:1.
- *Effect size* (*R*) sensitive to both magnitude of effect on at least one discriminating variable and the noise ratio (i.e., ratio of nondiscriminating variables to true discriminators); noise ratio relatively more important.
- *Data standardization* (col z-score) has moderate effect on effect size (*R*).

Tests of Among-Group Differences

Variations on a Good Theme

- Blocked MRPP (MRBP) – For randomized blocked designs, paired-sample data, and simple repeated measures (haven't found R function; but BLOSSOM software can do it).
- Q_b method – for partitioning variance in the distance matrix into sums of squares for multiple factors, including interactions (limited to euclidean distance metric).
- NPMANOVA – for non-euclidean distance measures in multifactor designs, including nested and factorial designs. [adonis(vegan)]

Tests of Among-Group Differences

Considerations with MRPP & Related Methods

- MRPP and related methods (ANOSIM, MRBP, Q_b, NPMANOVA) test for differences among groups in either *location* (differences in mean) or *spread* (differences in within-group distance). That is, they may find a significant difference between groups simply because one of the groups has greater dissimilarities (dispersion) among its sampling units.
- Differences in spread ('heterogeneity of dispersions') can be either an asset or an assumption that must be met.

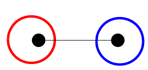
Tests of Among-Group Differences

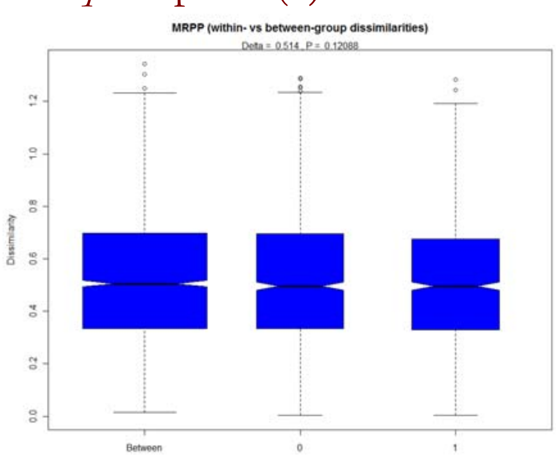
Considerations with MRPP & ANOSIM

Simulated Data Sets

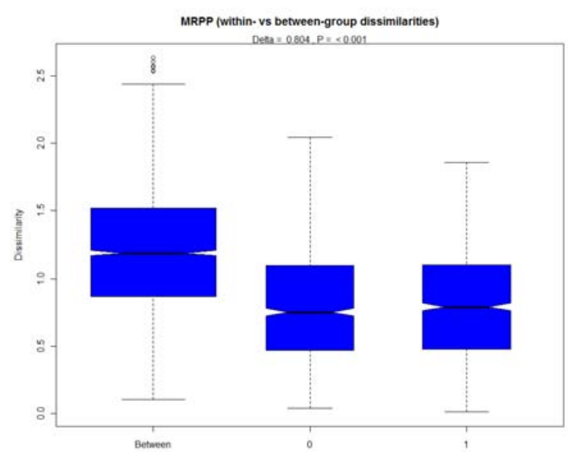
Equal location(e)
Equal spread(d)

Unequal location(e)
Equal spread(d)





n=100, p=1, e=0, d=1



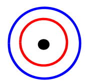
n=100, p=1, e=1, d=1

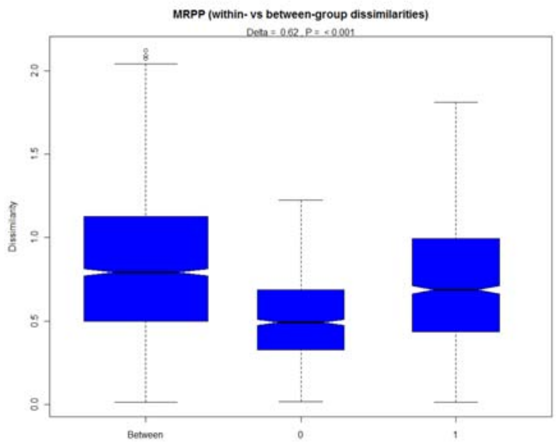
Tests of Among-Group Differences

Considerations with MRPP & ANOSIM

Simulated Data Sets

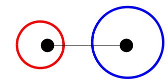
Equal location(e)
Unequal spread(d)

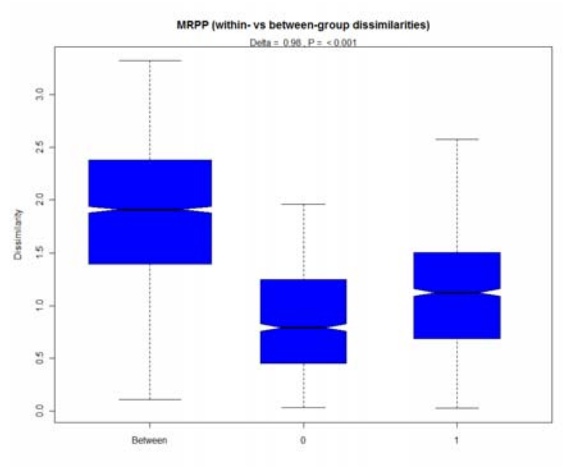




n=100, p=1, e=0, d=2

Unequal location(e)
Unequal spread(d)





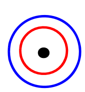
n=100, p=1, e=1, d=2

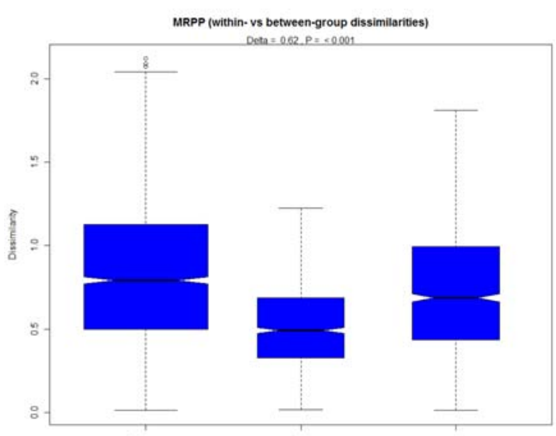
Tests of Among-Group Differences

Considerations with MRPP & ANOSIM

Simulated Data Sets

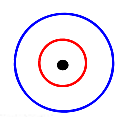
Equal location(e)
Unequal spread(d)

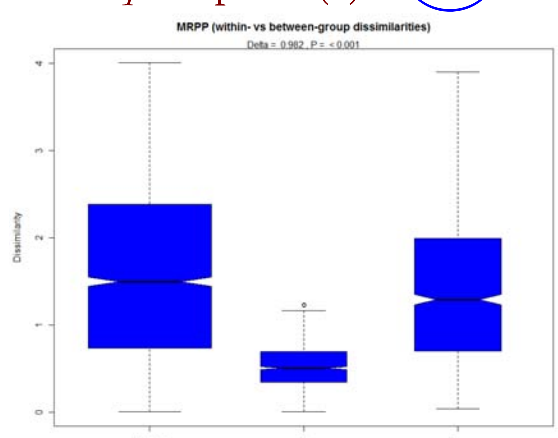




n=100, p=1, e=0, d=2

Equal location(e)
Unequal spread(d)





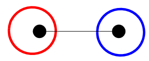
n=100, p=1, e=0, d=4

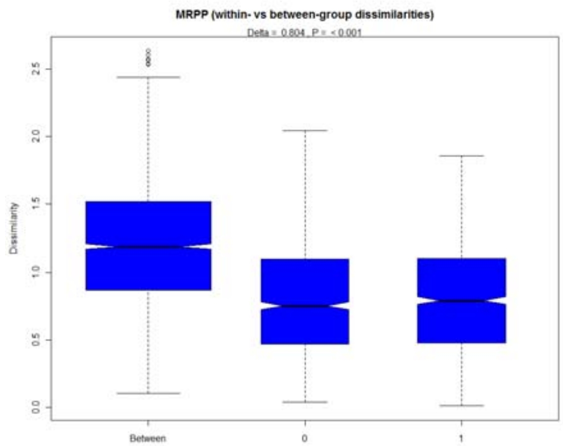
Tests of Among-Group Differences

Considerations with MRPP & ANOSIM

Simulated Data Sets

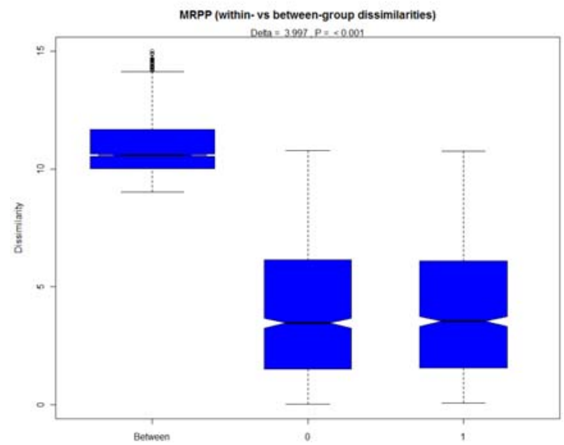
Unequal location(e)
Equal spread(d)





n=100, p=1, e=1, d=1

Unequal location(e)
Equal spread(d)



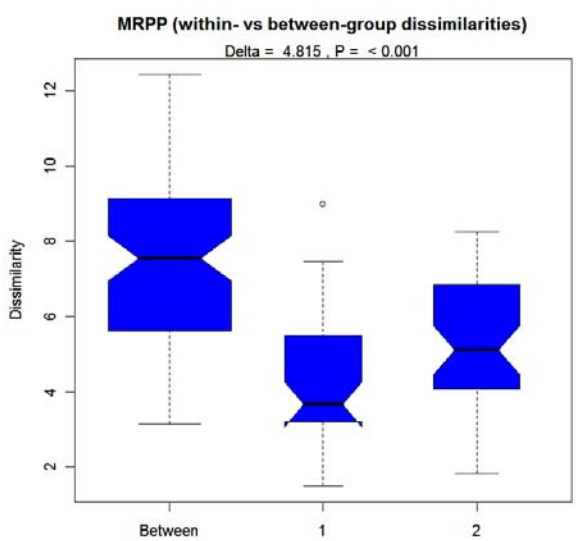
n=100, p=1, e=10, d=1

Tests of Among-Group Differences

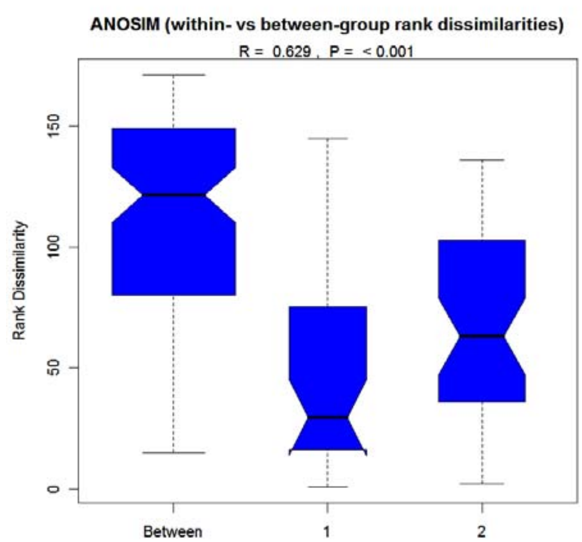
Considerations with MRPP & ANOSIM

Real data set example

MRPP



ANOSIM



Tests of Among-Group Differences

Assumptions of MRPP & Related Methods

- *Independent samples* -- Usual problems associated with pseudoreplication, subsampling and repeated measures apply here.
- Chosen *distance measure* adequately represents the variation of interest in the data.
- The *relative weighting of variables* has been controlled prior to calculating distance, such that the weighting of variables is appropriate for the ecological question at hand.

Tests of Among-Group Differences

Mantel's Test (MANTEL)

- Mantel statistic tests for differences between two distance matrices (e.g., between ecological and geographic distances between points), but can also be used to test for differences among groups.

- Calculate *dissimilarity* matrices.
- Calculate *test statistic* Z.

$$\boxed{z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_{ij} y_{ij}}$$

Hadamard Product

$$\begin{bmatrix} \text{Y} \\ \textit{Ecological} \\ \textit{distance} \end{bmatrix} \begin{bmatrix} \text{X} \\ \textit{Contrast matrix for} \\ \textit{groups} \end{bmatrix}$$

0 = Same groups

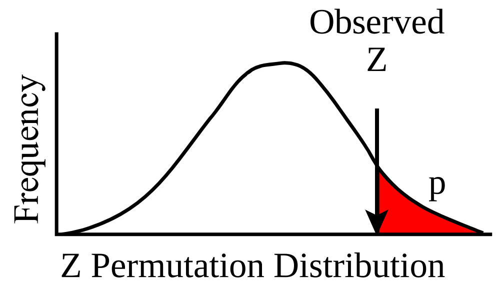
1 = Different groups

$$\begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & 1 & 0 & & \\ 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Tests of Among-Group Differences

Mantel's Test (MANTEL)

- Determine the *probability* of *a Z* this large or larger through Monte carlo permutations.
  - Permutations involve randomly shuffling one of the two distance matrices.
  - The significance test is simply the fraction of permuted *Z*'s that are greater than the observed Z.



Note, if ecological distances are rank transformed, this test is the same as ANOSIM and similar to rank-transformed MRPP.

Tests of Among-Group Differences

Mantel's Test (MANTEL)

- Determine the strength of the relationship between the two matrices by computing their correlation, r (*standardized Mantel statistic*).

$$r = \frac{\sum_i \sum_j \left[\frac{(x_{ij} - \bar{x})}{s_x}\right]\left[\frac{(y_{ij} - \bar{y})}{s_y}\right]}{n - 1}$$

$r > 0$ indicates that the average within-group distance between samples is less than the average overall distance between samples; the larger the r, the more tightly clustered the groups are.

$r < 0$ numerically possible, but highly unlikely in ecology.

Tests of Among-Group Differences

Mantel's Test (MANTEL)

2-group bird guilds example

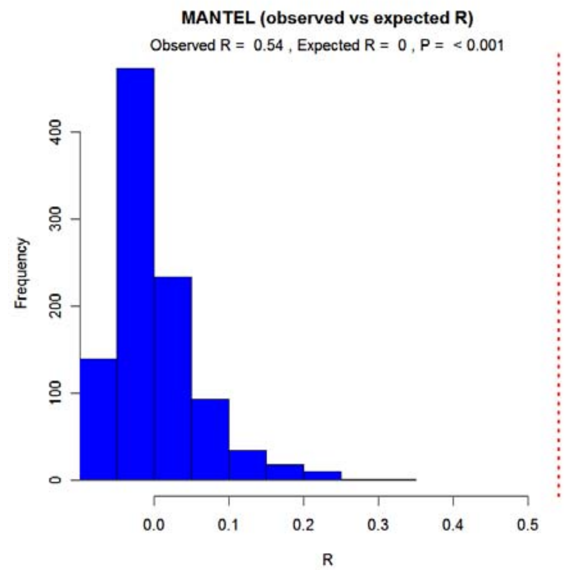
```
Mantel statistic based on Pearson's produ

Call:
mantel2(xdis = y.eucl, ydis = contrast)

Mantel statistic r: 0.5404
      Significance: < 0.001

Empirical upper confidence limits of r:
  90%   95% 97.5%   99%
0.084 0.122 0.176 0.211

Based on 1000 permutations
```



- Conclude that two clusters differ *significantly* in terms of the measured habitat variables.

Tests of Among-Group Differences

Mantel's Test (MANTEL)

Variations on a Really Good Theme

- Because Mantel's test is merely a correlation between distance matrices and the distance matrices can be variously defined, the test can assume a variety of forms as special cases.
- These are, in fact, variants of the same case but are interpreted somewhat differently. There are at least six variants.

[Y Dissimilarity matrix] [X Dissimilarity matrix] [Z Dissimilarity matrix]

Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 1. Simple Mantel's Test on Geographic Distance.*

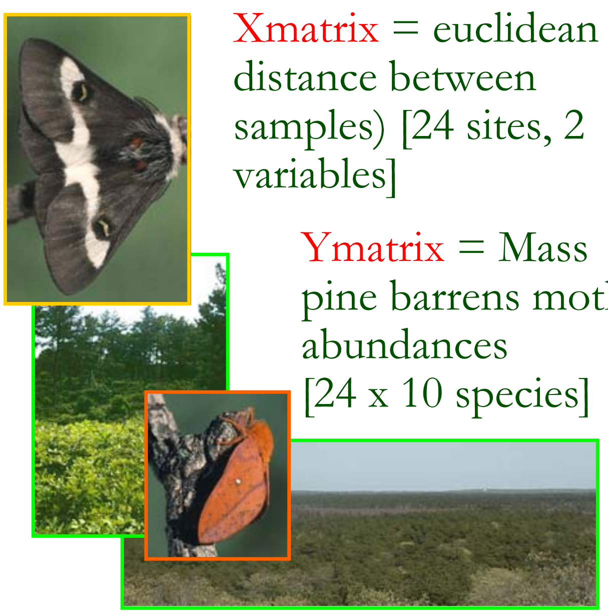
  If the dependent distance matrix is species similarity and the predictor matrix is geographic distance ("spatial dissimilarity"), the research question is "Are samples that are close together also compositionally similar?" This is equivalent to testing for overall autocorrelation in the dependent matrix (*i.e.*, averaged over all distances).

$$\left[\ \textit{Species Dissimilarity}\ \right] \left[\ \textit{Geographic Distance}\ \right]$$

Tests of Among-Group Differences

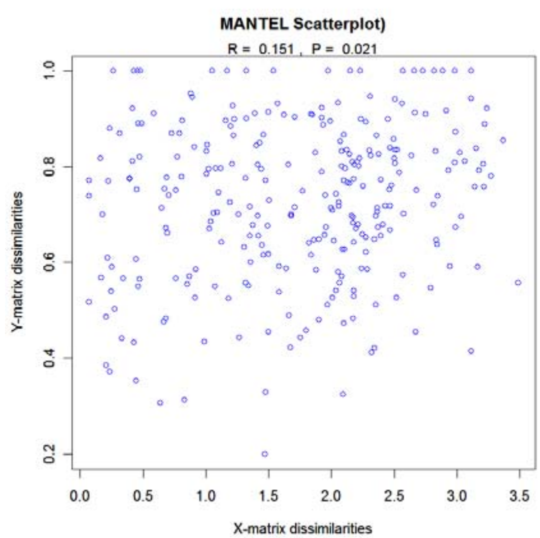
Mantel's Test (MANTEL)

- *Case 1. Simple Mantel's Test on Geographic Distance.*



Xmatrix = euclidean distance between samples) [24 sites, 2 variables]

Ymatrix = Mass pine barrens moth abundances [24 x 10 species]



Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 2. Simple Mantel's Test on a Predictor Matrix.*

If the dependent matrix is again species similarity and the predictor matrix is a dissimilarity matrix based on a set of environmental variables, then the simple test is for correlation between the two matrices. Such correlation would indicate that locations that are similar environmentally tend to be similar compositionally. This, of course, is one of the fundamental questions in ecology.

$$\left[\begin{array}{c}\textit{Species}\\ \textit{Dissimilarity}\end{array}\right]\left[\begin{array}{c}\textit{Environmental}\\ \textit{Dissimilarity}\end{array}\right]$$

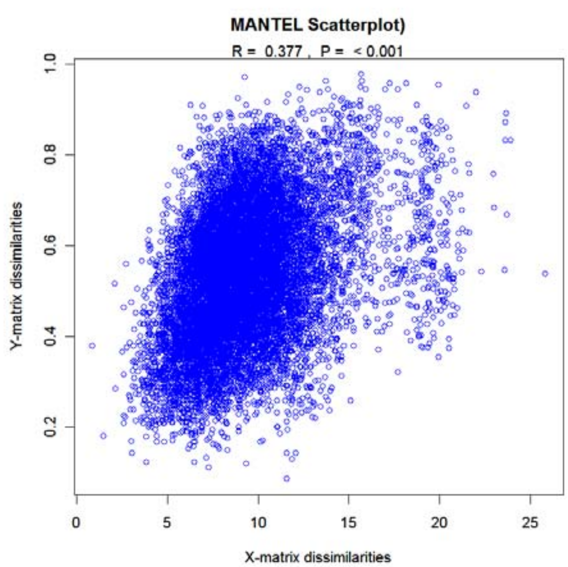
Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 2. Simple Mantel's Test on a Predictor Matrix.*

Xmatrix =
Oregon riparian environment (geomorphology, landscape composition and configuration) [164 sites, 48 variables]

Ymatrix =
Oregon riparian birds abundances [164 sites, 49 species]



Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 3. Simple Mantel's Test between an Observed Matrix and One Posed by a Model.*

  As a formal hypothesis test, Mantel's test can be used to compare an observed dissimilarity matrix to one posed by a conceptual or numerical model. Here, the model is provided as a user-provided matrix of similarities or distances, and the test is to summarize the strength of the correspondence between the two matrices. The model distance matrix might be provided as a simple binary matrix of 0's and 1's, or a matrix derived from a more complicated model.

[*Observed Dissimilarity*] [*Model Dissimilarity*]

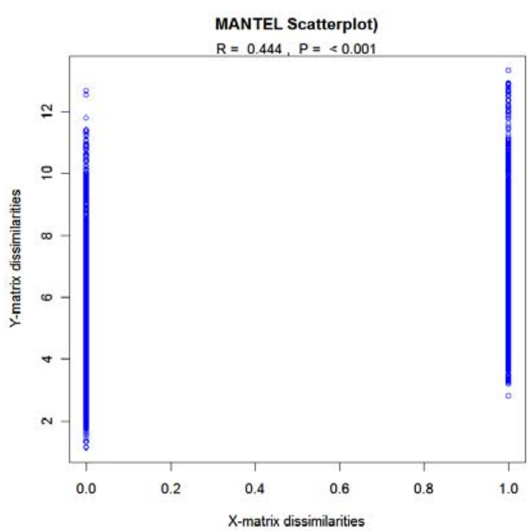
Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 3. Simple Mantel's Test between an Observed Matrix and One Posed by a Model.*

Xmatrix =
Hammond's flycatcher presence/absence in Oregon Coast Range
[96 sites, 1 indicator variable]

Ymatrix =
Oregon habitat variables
[96 sites, 20 habitat variables]



Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 4. The Mantel Correlogram.*

  A special case of case 1 (above) is to partition or subset the analysis into a series of discrete distance class. That is, a first distance matrix is evaluated for all pairs of points within the first distance class; then a second matrix is scored for all pairs of points within the second distance interval, and so on. The result of this analysis is a *Mantel's correlogram*, completely analogous to an autocorrelation function but performed on a (possibly multivariate) distance matrix.

$$\begin{bmatrix} \textit{Observed} \\ \textit{Dissimilarity} \end{bmatrix} \begin{bmatrix} \textit{Geographic} \\ \textit{Distance Class} \end{bmatrix}$$

Tests of Among-Group Differences

Mantel's Test (MANTEL)

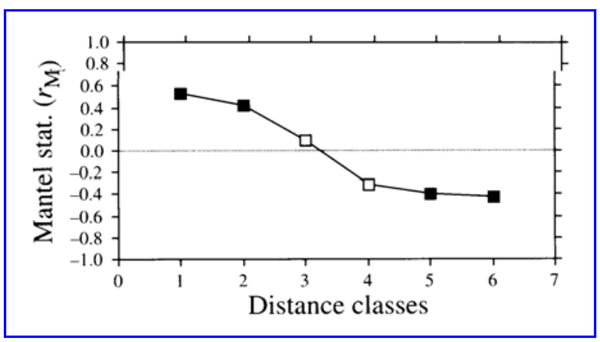
- *Case 4. The Mantel Correlogram.*

$$\begin{bmatrix} \text{X} \\ \textit{Ecological} \\ \textit{distance} \end{bmatrix} \begin{bmatrix} \text{Y} \\ \textit{Geographic} \\ \textit{distance} \end{bmatrix}$$

$$\begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 2 & 0 & 2 & & \\ 0 & 0 & 0 & 2 & \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix} r_{m2} \qquad \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & 1 & 0 & & \\ 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} r_{m1}$$

Etc.

Mantel correlogram



Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 5. Partial Mantel's Test on Three Distance Matrices.*

The idealized Mantel's test is a partial regression on three distance matrices. Here, the research question is, "How much of the variability in the dependent matrix is explained by the independent matrix after removing the effects of a third constraining matrix?" The analysis in this case is partial regression, and both partial correlation (or regression) coefficients are of interest: rYX|Z and rYZ|X.

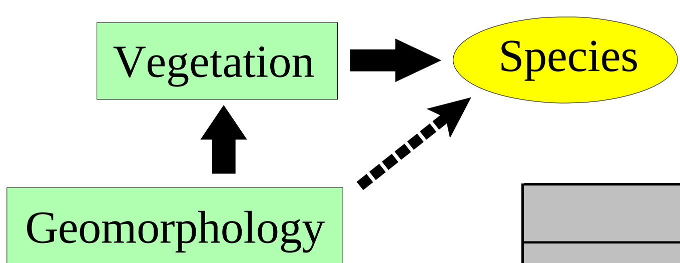
$$\begin{bmatrix} Y \\ \textit{Dissimilarity} \\ \textit{matrix} \end{bmatrix} = \begin{bmatrix} X \\ \textit{Dissimilarity} \\ \textit{matrix} \end{bmatrix} \Bigg| \begin{bmatrix} Z \\ \textit{Dissimilarity} \\ \textit{matrix} \end{bmatrix}$$

Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 5. Partial Mantel's Test on Three Distance Matrices.*

Path Diagram (Oregon streamside bird communities)



Simple Mantel r's

Partial Mantel r's

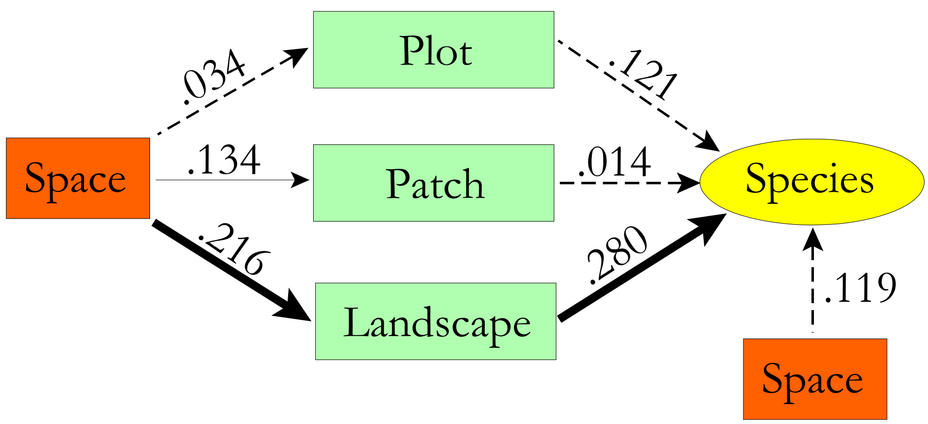
<table>
<tr><th></th><th>Species</th><th>Geomorphology</th><th>Vegetation</th></tr>
<tr><td>Species</td><td>-</td><td>Geo-Spec
.266 (p<.001)</td><td>Veg-Spec
.326 (p<.001)</td></tr>
<tr><td>Geomorphology</td><td>Spec-Geo|Veg
.211 (p<.001)</td><td>-</td><td>Veg-Geo
.219 (p=.003)</td></tr>
<tr><td>Vegetation</td><td>Spec-Veg|Geo
.285 (p<.001)</td><td>Geo-Veg|Spec
.146 (p=.019)</td><td>-</td></tr>
</table>

Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 5. Partial Mantel's Test on Multiple Distance Matrices.*

Path Diagram (Mass. pine barrens moth community)



Tests of Among-Group Differences

Mantel's Test (MANTEL)

- *Case 6. Partial Mantel's on Multiple Predictor Variables.*

Often, knowing that the environment has some relationship with the dependent variable of interest is not sufficiently satisfying: we wish to know which variables are actually related to the dependent variable. The logical extension of Mantel's test is multiple regression, in which the predictor variables are entered into the analysis as individual distance matrices. As a partial regression technique, Mantel's test provides not only an overall test for the relationships among distance matrices, but also tests the contribution of each predictor variable for its pure partial effect on the dependent variable.