

Utah State University

DigitalCommons@USU

All Graduate Plan B and other Reports

Graduate Studies

5-2017

A Comparison of Statistical Methods RELating Pairwise Distance to a Binary Subject-Level Covariate

Rachael Stone

Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/gradreports>



Part of the [Applied Statistics Commons](#), [Biostatistics Commons](#), and the [Statistical Models Commons](#)

Recommended Citation

Stone, Rachael, "A Comparison of Statistical Methods RELating Pairwise Distance to a Binary Subject-Level Covariate" (2017). *All Graduate Plan B and other Reports*. 931.

<https://digitalcommons.usu.edu/gradreports/931>

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



A COMPARISON OF STATISTICAL METHODS RELATING PAIRWISE DISTANCE
TO A BINARY SUBJECT-LEVEL COVARIATE

By

Rachael L. Stone

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Statistics

Approved:

Dr. John R. Stevens
Major Professor

Dr. Yan Sun
Committee Member

Dr. Chris Corcoran
Committee Member

Utah State University
Logan, Utah
2017

Copyright © Rachael L. Stone 2017
All Rights Reserved

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor, Dr. John Stevens, for his continuous support of my graduate level studies, as well as my undergraduate studies. His patience, motivation, and knowledge have been deeply appreciated. His guidance helped me during the time of research and writing of this thesis

Besides my advisor, I would like to thank the other members of my thesis committee, Department Head and Professor Chris Corcoran and Doctor Yan Sun.

I thank my husband, Jacob, for his love and support in my thesis work, as well as every other aspect of my life.

My parents Richard and Sherilyn deserve much appreciation. They have always been greatly supportive and emphasized education throughout my life. Special mentions to my brothers: Matthew, Michael, and William; along with my other family members: Tracy, Lucy and Emmett and the Stones. My family has shown me great love and encouragement during my schooling.

My sincere thanks also go to Pluralsight, a company I have thoroughly enjoyed working for and where I have encountered wonderful knowledgeable co-workers, who have taught me so much. The Revenue Platform Team deserves an extra special thank you. They have provided me with wonderful learning opportunities, helped me grow my career-related skill set, and helped me implement my knowledge from the M.S. program.

Lastly, I would like to thank Jerry Hintze for generously giving his support and encouragement in my studying of statistics. He has been a strong and positive role model as a statistician, friend, and overall person.

ABSTRACT

A Comparison of Statistical Methods Relating Pairwise Distance to a Binary Subject-Level Covariate

by

Rachael L. Stone

Utah State University, 2017

Major Professor: Dr. John R. Stevens
Department: Mathematics and Statistics

A community ecologist provided a motivating data set involving a certain animal species with two behavior groups, along with a pairwise genetic distance matrix among individuals. Many community ecologists have analyzed similar data sets with a method known as the Hopkins method, testing for an association between the subject-level covariate (behavior group) and the pairwise distance. This community ecologist wanted to know if they used the Hopkins method, would their results be meaningful? Their question inspired this thesis work, where a different data set was used for confidentiality reasons. Multiple methods (Hopkins method, ADONIS, ANOSIM, and Distance Regression) were used to analyze the distance matrix for association with a binary covariate of interest. To compare the performance of the Hopkins method with the performance of the remaining, more established methods, a simulation was run. The results of the simulation indicate that ADONIS, ANOSIM, and distance regression would all be preferable to the Hopkins method.

CONTENTS

ACKNOWLEDGEMENTS.....	3
ABSTRACT.....	5
CONTENTS.....	6
LIST OF FIGURES.....	7
LIST OF TABLES.....	8
BACKGROUND: MOTIVATIONAL STUDY	9
BACKGROUND: THE DUNE MEADOW VEGETATION STUDY	10
METHODS.....	14
DISCUSSION.....	19
REFERENCES.....	22
APPENDIX A – R SOURCE CODE.....	24

LIST OF FIGURES

Figure		Page
1	Simulated datasets for different values of δ	16
2	Each Methods' Power as δ Increases	18
3(a)	Heatmap of covariate A1 and between-site distance.....	20
3(b)	Heatmap of covariate Moisture and between-site distance.....	20
3(c)	Heatmap of covariate Manure and between-site distance.....	20

LIST OF TABLES

Table		Page
1	Vegetation Species Recorded in Dune Meadow Vegetation Study.....	11
2	Van der Maarel Variables used in Vegetation Measurements.....	11
3	Variables in <i>dune.env</i> and Their Binary Classification.....	12
4	Comparison of reported Type I Error Rates.....	17

BACKGROUND: MOTIVATING STUDY

Some time ago, a community ecologist approached Dr. Stevens (major professor to the author of this MS report) about a set of data being used in a study of their own. This data set referred to an animal species and a difference in some of this animal's behavior (we will refer to these behaviors as behavior 1 and behavior 2). A genetic distance matrix and a binary subject-level covariate were computed by this community ecologist, who wanted to evaluate the distance matrix differences between these 2 behavior groups.

Other community ecologists who have had similar datasets have used the Hopkins method (to be discussed in the next section) to relate the genetic distance matrix to a binary subject-level covariate. The community ecologist wanted to know: if the Hopkins method was used in the analysis of their data, would it result in meaningful calculations, or was there a better way to analyze the data?

Because this community ecologist's data are privately owned and not available for public distribution, this report uses the dataset *dune* (to be discussed in the next section) with the intention of evaluating the performance of several possible methods that could be used to test for significant associations between a pairwise distance matrix and a binary subject-level covariate.

BACKGROUND: THE DUNE MEADOW VEGETATION STUDY

The Dune Meadow Vegetation Study was conducted on the Dutch island of Terschelling (Batterink and Wiffels, 1983). The objective of this study was to detect a possible relation between the vegetation and the management of dune meadow on the island. Data collection was done by the Braun-Blanquet method; the data are recorded per the ordinal scale of van der Maarel (1979). This dataset was then used as a training example in Jongman et al. (1987) and are part of the *vegan* package in R (Oksanen 2017).

This dataset contains 20 plots of $2 \times 2 \text{ m}^2$ with recorded environmental variables of interest: A1 (a measurement of soil thickness), moisture, and manure. Each of the 20 plots were used as a sample area of a larger parcel of land and used to represent the corresponding larger parcel. There was a total of 80 parcels of land but only 20 were selected for inclusion in the data set *dune*. For each of these 20 different plots there was a quantified abundance of 30 different species of vegetation; this is the data contained in the *dune* data frame found in R. The 30-different species are identified in Table 1 (Jongman 1987).

These species of vegetation had their abundance quantified per the ordinal scale of van der Maarel (1979) and the Braun-Blanquet Method. The scale of van der Maarel uses specific variables (summarized in Table 2) in its quantified values while the Braun-Blanquet transformed these values to a cover-abundance scale on a scale of 0 to 9. Zero represents no vegetation while 9 represents a high density of vegetation (Jongman 1987) for the given species.

Table 1- List of the Types of Vegetation Species Recorded in The Dune Meadow Vegetation Study (Jongman 1987)

Vegetation Species		
Achimill	Eleopalu	Ranuflam
Agrostol	Elymrepe	Rumeacet
Airaprae	Empenigr	Sagiproc
Alopgeni	Hyporadi	Salirepe
Anthodor	Juncarti	Scorautu
Bellpere	Juncbufo	Trifprat
Bromhord	Lolipere	Trifrepe
Chnalbu	Planlac	Vicilath
Cirsarve	Poaprat	Bracruta
Comapalu	Poatriv	Callcusp

Table 2 – The specific variables used to calculate measurements in the scale of van der Maarel (van der Maarel 1979)

Environmental Variables	
Variable	Definition
Abundance	the number of individuals on the sample plot
Frequency	the number of times a species occurs in a sample plot
Cover	the estimated area that a species covers
cover-abundance	combined parameter of cover and abundance
basal-area	area outline of a plant near the surface
Phytomass	measure per species, their performance in time series

The *dune* meadow vegetation data has a corresponding environmental data frame *dune.env* within R that contains the environmental data for each of the 20 plots of land. For the purposes of this report, the variables are treated as binary. A brief description of how the variables were converted to a binary form is found in the “Binary Classification” column in the Table 3.

Table 3 – Variables found in the data frame *dune.env* within R (Oksanen 2017)

Environmental Variables in <i>dune.env</i> Dataset		
Name of variable	Description	Binary Classification
A1	Thickness of the A1 horizon in cm	A soils thickness of A1 ≤ 4 , A1 is classified as “Low”. A soils thickness of A1 >4 , is classified as “High”. 9 observations are classified as “Low”. 11 observations are classified as “High”.
Moisture	Moisture content of the soil (scale of 1-5)	A moisture content of 1 or 2 is categorized as “Low”. A moisture content of 4 or 5 is categorized as “High”. (No observations had a moisture content of 3.) 10 observations are classified as “Low”. 10 observations are classified as “High”.
Manure	Quantity of manure applied (scale of 0-4)	An observation is considered “Low” if manure is classified as a 0 or 1. An observation is “High” if manure is classified as a 2, 3 or 4. 9 observations are classified as “Low”. 11 observations are classified as “High”.

To compare statistical methods relating pairwise distance to a subject-level covariate, a pairwise distance matrix was needed for the dune dataset and the 20 plots of interest. This needed transformation was performed by the *vegdist* function from the *vegan* package (Oksanen 2017). This function is commonly used by community ecologists. Specifically, this function took each of the 20 plots and the corresponding Braun-Blanquet abundance scores found in the *dune* object for each vegetation species, and calculated pairwise plot dissimilarity indices popular with community ecologists. The collection of dissimilarity indices provided a 20 x 20 pairwise distance matrix which quantifies the dissimilarities of each plot's vegetation density across all species.

METHODS

Multiple methods were considered to analyze the distance matrix of interest (provided by the *dune* dataset) for association with a binary covariate of interest: the Hopkins method (Hopkins 2013), ADONIS (Anderson 2001), ANOSIM (Clarke 1993), and Distance Regression (Zapala and Schork 2006), all briefly summarized below.

The Hopkins method (Hopkins 2013) tests whether the average distance among the N_{high} observations with the same “High” level in the covariate of interest is significantly different from the average distance among all observations. This method performs bootstrap resampling of N_{high} sites from the total number of sites, and averages the values found within their distance matrix. Noting the number of times a bootstrapped mean distance was less than or equal to the mean distance among the observed N_{high} sites, a p-value was calculated by the proportion of times the bootstrapped mean distance were at least as extreme as (i.e., less than or equal to) the mean distance among the original N_{high} sites (Hopkins 2013).

Another method used for this type of comparison is ADONIS (Anderson 2001). It is a Permutational Multivariate Analysis of Variance Using Distance Matrices, or an Analysis of variance using distance matrices. ADONIS is also used for partitioning distance matrices among sources of variation, such as the two levels of the binary covariates of interest: low and high. It is also used for fitting linear models to distance matrices. It uses a pseudo-F ratio to calculate a p-value, and is implemented in the *adonis* function of the R package *vegan* (Oksanen 2017). The name ADONIS is a letter-swapped alteration on its original name of ANODIS (for Analysis of Dissimilarities); the alteration was made to avoid confusion with the fact that ANOSIM (described briefly below) also handles Analysis of Dissimilarities (Oksanen 2008).

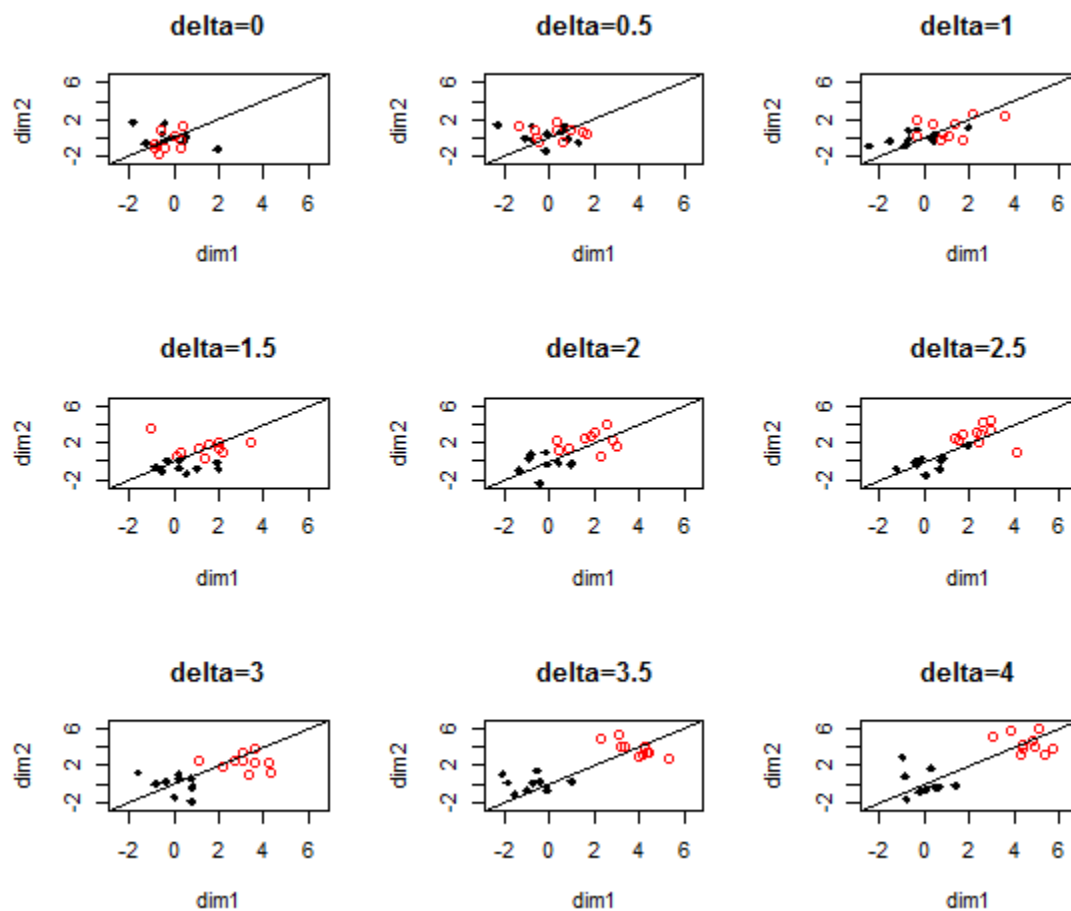
ANOSIM stands for Analysis of Similarities (Clarke 1993). This method tends to have similar results as ADONIS but is included for verification. The help file for the *anosim* function in the R package *vegan* (Oksanen 2017) states that this method “provides a way to test statistically whether there is a significant difference between two or more groups of sampling units” in terms of their similarities, dissimilarities, or distances. Within the dune example the plots of land would be our sampling units. Permuting the distance matrix’s “group membership” (or the vector of the sites’ covariates levels) many times is done to obtain the null distribution of the R statistic, which is the difference of mean ranks between groups and within groups. This calculated R, when compared to the null distribution, provides a decision of statistical significance (Oksanen 2008).

Distance regression (Zapala and Schork 2006) can be performed using the *dr* function of the *AssocTests* package in R (Wang *et al.* 2015). This is a distance regression that is used to detect the association between a distance matrix and some independent variants of interest (such as binary covariates in the *dune* example). The significance of a pseudo F statistic is assessed using a Monte Carlo permutation approach, returning a p-value.

To compare the performance of the Hopkins method with the performance of the more established ADONIS, ANOSIM, and distance regression methods, a simulation was run. This simulation called for 20 hypothetical sites, the same number of sites in the *dune* dataset. For each hypothetical site, there is a corresponding binary covariate with half of the sites at level 0 (low) and half at level 1 (high). Bivariate normal data was simulated for the level 0 and level 1 groups, with the level 0 group always centered at the origin (0,0), and the level 1 group centered at (δ , δ). The δ value was originally 0, and then progressively increased to $\delta=4$ by steps of 0.5. For each simulated data set, the Euclidean distance (in this two-dimensional space, for convenience) was

calculated between each pair of sites. The progression of this simulation is demonstrated visually in Figure 1.

Figure 1- Visualizations of the simulated datasets for different values of δ . The solid dots represent when level = 0 (or low) while the hollow circles represent when level = 1 (or high).



This simulation was run 1000 times, with 1000 bootstrap samples for methods employing resampling, and the δ values range from 0 to 4. For each simulated data set, Euclidean distance (in the 2-dimensional space visualized in Figure 1) was used to define a 20 x 20 pairwise

distance matrix among the 20 observations. As δ increases, the overall distance between the binary covariate's levels also increases (Figure 1). This simulation was repeatedly testing the 20 x 20 distance matrices' association with the binary covariate (group level 0 or group level 1), using the four methods summarized above (Hopkins, ADONIS, ANOSIM, and Distance Regression). The run time for this simulation was 5 hours and 34 minutes.

The proportion of all simulations giving a significant result when $\delta = 0$ resulted in a Type I error rate estimate, or the probability of calling significant association when there is no association, for each method. From this simulation, the power (or the probability of determining a significant association when there is a non-zero association) may also be estimated by taking the proportion of all simulations giving a significant result when $\delta > 0$, for each method. When looking at Type I error rates (reported in Table 4) across different methods, the Hopkins method suggested that it differed greatly from the other three methods. It had a *much* lower Type I error rate suggesting that it is overly conservative.

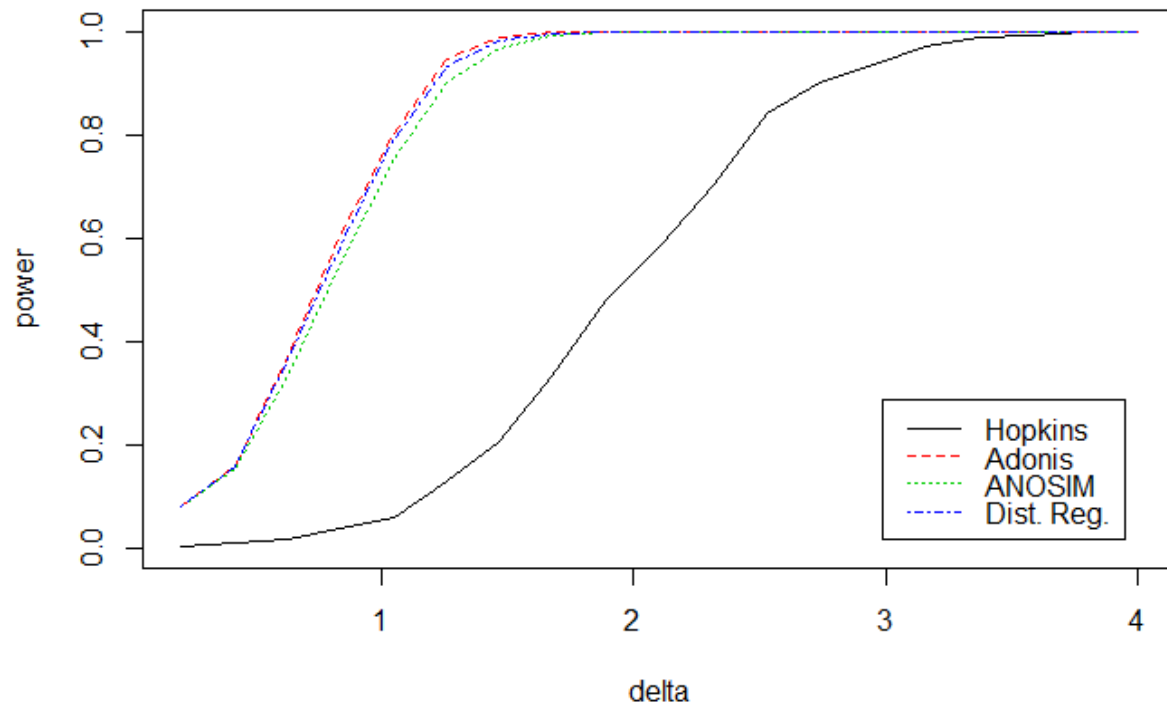
Table 4 – Comparison of reported Type I Error Rates

Type I Error Rate				
Method	Hopkins	ADONIS	ANOSIM	Distance Regression
Type I Error	0.005	0.049	0.048	0.056

The Hopkins method differs for another reason; it has less power than the other methods (see Figure 2). ADONIS, ANOSIM and the distance regression methods are all like one another; they perform almost equally well. Although, ADONIS seems to, just slightly, outperform

ANOSIM and Distance Regression. Figure 2 visualizes the difference in power between methods.

Figure 2- Visual Representation of each Methods' Power as δ increases



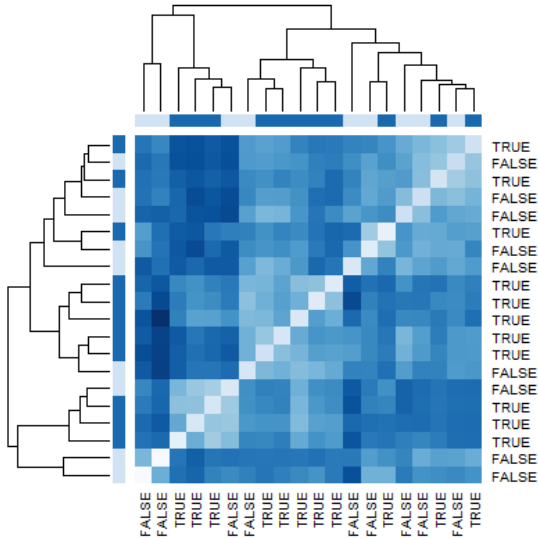
DISCUSSION

When comparing these methods side by side (especially in terms of type I error rates in Table 4 and power in Figure 2), it is clear, the Hopkins method is not preferred. The Hopkins method is out performed by the remaining methods. ADONIS, ANOSIM and Distance Regression are very similar and provide almost identical results in terms of power and Type I error rate. If a comparison of distance matrices is being performed for this type of study, then ADONIS, ANOSIM and Distance Regression are all acceptable with ADONIS being slightly more recommended due to it being slightly more powerful.

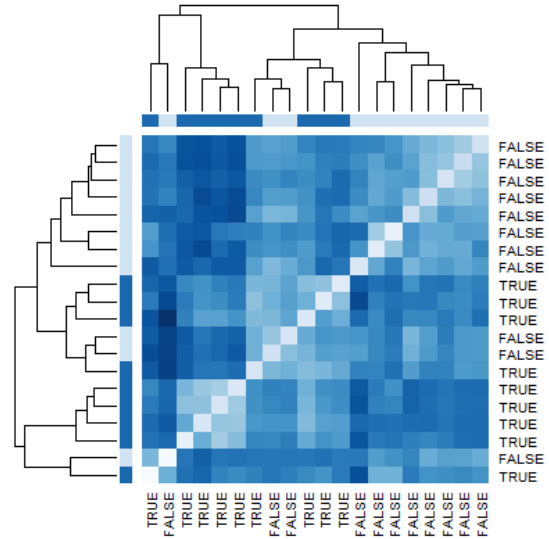
With ADONIS being the best-performing method, it was then applied it to the *dune* data set. It was applied to the 3 binary covariates (A1, Moisture, and Manure). When applying ADONIS to the variable A1, a p-value of 0.098 was calculated. With this p-value, it is concluded that there is not a significant association between A1 (high/low) and between-site distance at α -level equal to 0.05 (Figure 3a is a visualization of the association between A1 and between-site distance.) A p-value of 0.001 was calculated for the variable Moisture with a conclusion that there is an association between Moisture (high/low) and between-site distance at α -level equal to 0.05 (Figure 3b is a visualization of the association between Moisture and between-site distance.) A p-value of 0.253 was calculated for the variable Manure with a conclusion that there is not an association between (high/low) and between-site distance at α -level equal to 0.05 (Figure 3c is a visualization of the association between Manure and between-site distance.)

Figure 3 – A heatmap visualization of the dune pairwise between-site distance matrix with column- and row-side colors and labels corresponding to levels of covariates (a) A1 high, (b) Moisture high, and (c) Manure high. The title of each Figure panel includes the ADONIS p-value for the covariate’s association with between-site distance.

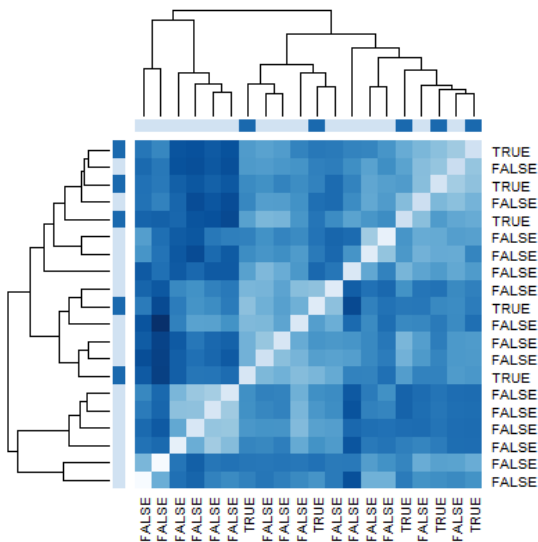
(a) A1 high, p-value 0.089



(b) Moisture high, p-value 0.001



(c) Manure high, p-value 0.253



Of the three binary subject-level covariates considered in the *dune* data, it appears that only Moisture is significantly associated with the between-site similarities in vegetation density. Figure 3b helps visualize this – the generally-similar set of eight sites represented by the light colored block in the upper-right of the heatmap (and the right third of the dendrogram at top) all have the same FALSE value for the binary covariate Moisture high. In other words, these most-similar sites (in terms of vegetation density across the 30 recorded species) have a low level of Moisture. There is another smaller block of four generally-similar sites (in terms of vegetation density across the 30 recorded species) near the lower-left corner of the heatmap that all have a high level of Moisture. The generally-similar blocks of sites in Figures 3a and 3c apparently do not correspond sufficiently to the high/low levels of A1 or Manure, respectively, to cause ADONIS to return a significant result for those covariates.

After calculating comparisons of these methods and applying them to the dune dataset, it would not be suggested that the Hopkins method be applied to similar datasets, including the dataset provided by the community ecologist mentioned in the motivating study above. Instead, ADONIS, ANOSIM, and distance regression would all be improved methods compared to the use of the Hopkins method, particularly to achieve high statistical power and avoid an overly-conservative type I error rate when testing for association of a pairwise distance matrix with a binary subject-level covariate.

REFERENCES

- Anderson, M.J. (2001). *A new method for non-parametric multivariate analysis of variance*. *Austral Ecology*, 26: 32–46.
- Batterink M. and Wijffels G. (1983): *Een vergelijkend vegetatiekundig onderzoek naar de typologie en invloeden van het beheer van 1973 tot 1982 in de duinweilanden op Terschelling*. Report Agricultural University, Department of Vegetation Science, Plant Ecology and Weed Science, Wageningen.
- Buttigieg PL, Ramette A (2014). *A Guide to Statistical Analysis in Microbial Ecology: a community-focused, living review of multivariate data analyses*. *FEMS Microbiol Ecol*. 90: 543–550.
- Clarke, K. R. (1993). Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* 18, 117-143.
- Hopkins, J. B., III. (2013) *Use of genetics to investigate socially learned foraging behavior in free-ranging black bears*. *Journal of Mammalogy*. Oxford University Press.
- Jongman R.H.G, ter Braak C.J.F., and van Tongeren O.F.R. (1987). *Data Analysis in Community and Landscape Ecology*. Pudoc, Wageningen
- Oksanen J., Blanchet F. G., Friendly M., Kindt R., Legendre P., McGlenn D., Minchin P. R., O'Hara R. B., Simpson G. L., Solymos P., Stevens M. H. H., Szoecs E. and Wagner H. (2017). *vegan: Community Ecology Package*. R package version 2.4-2. <https://CRAN.R-project.org/package=vegan>
- Oksanen, J. (2008) [Vegan-devel] multiple regression on distance matrices. *R-Project*. Web. 13 Apr. 2017. <http://lists.r-forge.r-project.org/pipermail/vegan-devel/2008-October/000051.html>

Maarel, E. Van der (1979). *Transformation of cover-abundance values in phytosociology and its effects on community similarity*. SpringerLink. Kluwer Academic Publishers.

Wang L., Zhang W., Li Q. and Zhu W. (2015). AssocTests: Genetic Association Studies. R package version 0.0-3.
<https://CRAN.R-project.org/package=AssocTests>

Zapala, M.A. and N.J. Schork (2006). *Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables*. Proceedings of the National Academy of Sciences, USA, **103**:19430–19435.

APPENDIX A: R CODE

```
library(data.table)
library(vegan)
library(cluster)
library(AssocTests)
library(MASS)
library(sna)

data(dune)
data(dune.env)

duneNum = c(1:20)

duneMatrix <- as.matrix(vegdist(dune))

set.seed(123)

n <- 20 #total number of dunes
B <- 1000 #how many times i resample for bootstrap
nsim <- 1000

delta.vec <- round(seq(from=0,to=4,length.out=20),2)

p1 <- p2 <- p3 <- p4 <- matrix(ncol=length(delta.vec),nrow=nsim)

# Define binary variables
#A1
Alhigh <- dune.env$A1 > 4
# 10 False, 10 True

#Moisture
Moisturehigh=c(dune.env$Moisture == 4 | dune.env$Moisture == 5)
# 11 False, 9 True

#Manure
Manurehigh=c(dune.env$Manure == 2 | dune.env$Moisture == 3 |
             dune.env$Moisture ==4)
# 11 False, 9 True

par(mfrow=c(3,3))

#define methods 1-4

#Hopkins Approach
meth1 <- function(dX, fact)
{
```

```

meandX <- mean(as.matrix(dX)[fact==1, fact==1])

booteddX=rep(NA, B)

for (o in 1:B )
{
  dXSample <- sample(duneNum, sum(fact), replace=T)
  dXSampleMatrix=as.matrix(dX)[c(dXSample), c(dXSample)]
  booteddX[o] <- mean(dXSampleMatrix)
}

pvalue <- mean(booteddX <= meandX)
return(pvalue)
}

#ADONIS
meth2 <- function(dX, fact)
{
  fact <- adonis(as.matrix(dX)~fact, permutations = B)
  pvalue <- as.matrix(fact$aov.tab)[1,6]
  return(pvalue)
}

#ANOSIM
meth3 <- function(dX, fact)
{
  pvalue <- anosim(as.matrix(dX), fact, permutations = B)$signif
  return(pvalue)
}

#Distance Regression
meth4 <- function(dX, fact)
{
  #Assoc. Test
  x.mat <- cbind(rep(1,20), 1*fact)
  pvalue <- dr(as.matrix(dX), null.space=1, x.mat, permute=TRUE,
              n.MonterCarlo=B)$p.value
  return(pvalue)
}

print(date())

# Run simulations
for(i in 1:length(delta.vec))
{
  delta <- delta.vec[i]
  for(sim in 1:nsim)

```

```

{
  # Get distance matrix for simulated data
  X0 <- mvrnorm(n=10, mu=c(0,0), Sigma=diag(2))
  X1 <- mvrnorm(n=10, mu=c(delta,delta), Sigma=diag(2))

  X <- rbind(X0,X1)
  fact <- rep(c(0,1),each=10)
  dX <- dist(X)

  if(sim==1)
  {
    use.pch <- fact
    use.pch[fact==0] <- 18
    rr <- c(-2.5,2.5+max(delta.vec))
    plot(X, col=fact+1, pch=use.pch, xlab='dim1', ylab='dim2',
         main=paste('delta=',delta,sep=''),
         xlim=rr,ylim=rr)
    abline(0,1)
  }

  # Get p-values for the four methods

  p1[sim,i] <- meth1(dX,fact)
  p2[sim,i] <- meth2(dX,fact)
  p3[sim,i] <- meth3(dX,fact)
  p4[sim,i] <- meth4(dX,fact)
}
}

print(date())

par(mfrow=c(1,1))

## Summarize results
alpha <- .05

# When delta=0 (first column of p1, p2, p2, p4)
# get Type I error rate as pct sims giving p-value < alpha
err <- rep(NA,4)
err[1] <- mean(p1[,1] < alpha)
err[2] <- mean(p2[,1] < alpha)
err[3] <- mean(p3[,1] < alpha)
err[4] <- mean(p4[,1] < alpha)

err

# For each delta > 0 (columns 2+ of p1, p2, p3, p4)
# get Power for each method as pct sims giving p-value < alpha
power <- matrix(nrow=length(delta.vec)-1, ncol=4)

```

```

for(j in 1:length(delta.vec)-1)
{
  power[j,1] <- mean(p1[,j+1] < alpha)
  power[j,2] <- mean(p2[,j+1] < alpha)
  power[j,3] <- mean(p3[,j+1] < alpha)
  power[j,4] <- mean(p4[,j+1] < alpha)
}
row.names(power) <- delta.vec[-1]
colnames(power) <- c('Hopkins','Adonis','ANOSIM','Dist. Reg.')

#Visualize: plot pow1, pow2, pow3, pow4 (as lines)
matplot(rownames(power), power, type='l', xlab='delta',
        ylab='power', col=1:4)
legend('bottomright', inset=.05, legend=colnames(power),
       lty=1:4, horiz=FALSE, col=1:4)

#Apply ADONIS to Dune Dataset's binary variables
meth2(duneMatrix, Alhigh)
meth2(duneMatrix, Moisturehigh)
meth2(duneMatrix, Manurehigh)

#visualize the ADONIS results
heat <- function(dX, fact)
{
  colnames(duneMatrix) <- rownames(duneMatrix) <- fact
  bramp <- colorRampPalette(brewer.pal(n=9,"Blues"))(256)
  sc <- rep(bramp[50],length(fact))
  sc[fact] <- bramp[200]
  heatmap(duneMatrix, col=bramp, ColSideColors=sc, RowSideColors=sc)
}

#heatmaps of binary covariates
heat(duneMatrix, Alhigh)
heat(duneMatrix, Moisturehigh)
heat(duneMatrix, Manurehigh)

```