HW4

Kevin Mack

2/15/2016

Multivariate Modeling

Working with data about dune plants from Denmark

```
library(vegan)
```

```
## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.3-3
```

```
data(dune) #30 species each with 20 observations
data(dune.env) #data on the environemnt
?dune
names(dune)
```

```
##  [1] "Achimill" "Agrostol" "Airaprae" "Alopgeni" "Anthodor" "Bellpere"
##  [7] "Bromhord" "Chenalbu" "Cirsarve" "Comapalu" "Eleopalu" "Elymrepe"
## [13] "Empenigr" "Hyporadi" "Juncarti" "Juncbufo" "Lolipere" "Planlanc"
## [19] "Poaprat"  "Poatriv"  "Ranuflam" "Rumeacet" "Sagiproc" "Salirepe"
## [25] "Scorautu" "Trifprat" "Trifrepe" "Vicilath" "Bracruta" "Callcusp"
```

```
names(dune.env)
```

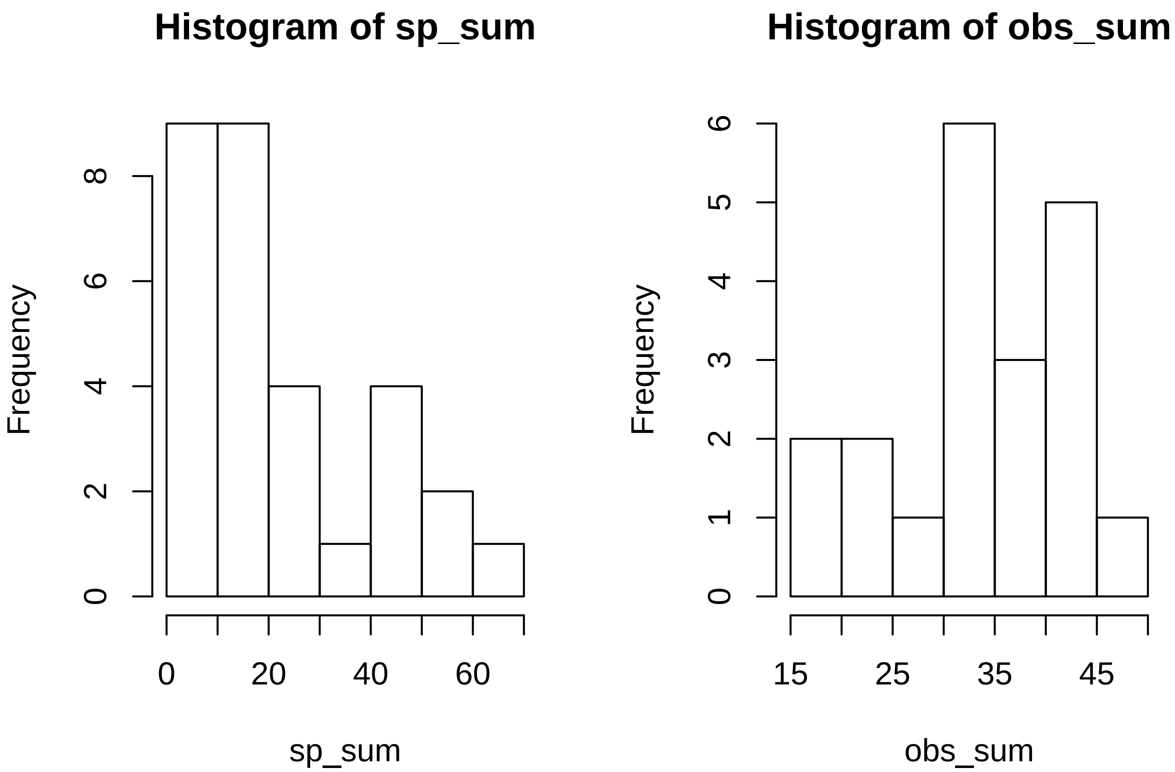
```
## [1] "A1"         "Moisture"   "Management" "Use"        "Manure"
```

```
summary(dune.env)
```

```
##        A1         Moisture Management       Use    Manure
##  Min.   : 2.800   1:7      BF:3       Hayfield:7   0:6
##  1st Qu.: 3.500   2:4      HF:5       Haypastu:8   1:3
##  Median : 4.200   4:2      NM:6       Pasture :5   2:4
##  Mean   : 4.850   5:7      SF:6                    3:4
##  3rd Qu.: 5.725                                    4:3
##  Max.   :11.500
```

1. We are interested in using indirect ordination methods (NMDS) to examine the role that moisture plays in dune plant communities. A quick visual summary indicates that some species are much more common than others, and that community assemblages vary.

```r
sp_sum = apply(dune, 2, sum)
obs_sum = apply(dune, 1, sum)
par(mfrow=c(1,2))
hist(sp_sum) #give me a histogram of the frequency each species occurs
#(ie, there are x species that occur 1000 times)
hist(obs_sum) #give me a histogram of the frequency of species in each observation/site?
```




```r
#(there are x sites that have 40 species in them)
par(mfrow=c(1,1))
```

Representing NMDS visually, we construct a plot that groups species and indicates environmental moisture levels. This plot indicates that generally, species assemblages are split based on moisture needs. It seems that species could be predicted based on high (4-5) and low (1-2) moisture levels. I would hypothesize that areas of dune with high moisture would host plant communities significantly different from those with low moisture.

```r
# Non-metric multidimenstional scaling (MDS) with default  Bray-Curtis.
dune_mds = (metaMDS(dune))
```

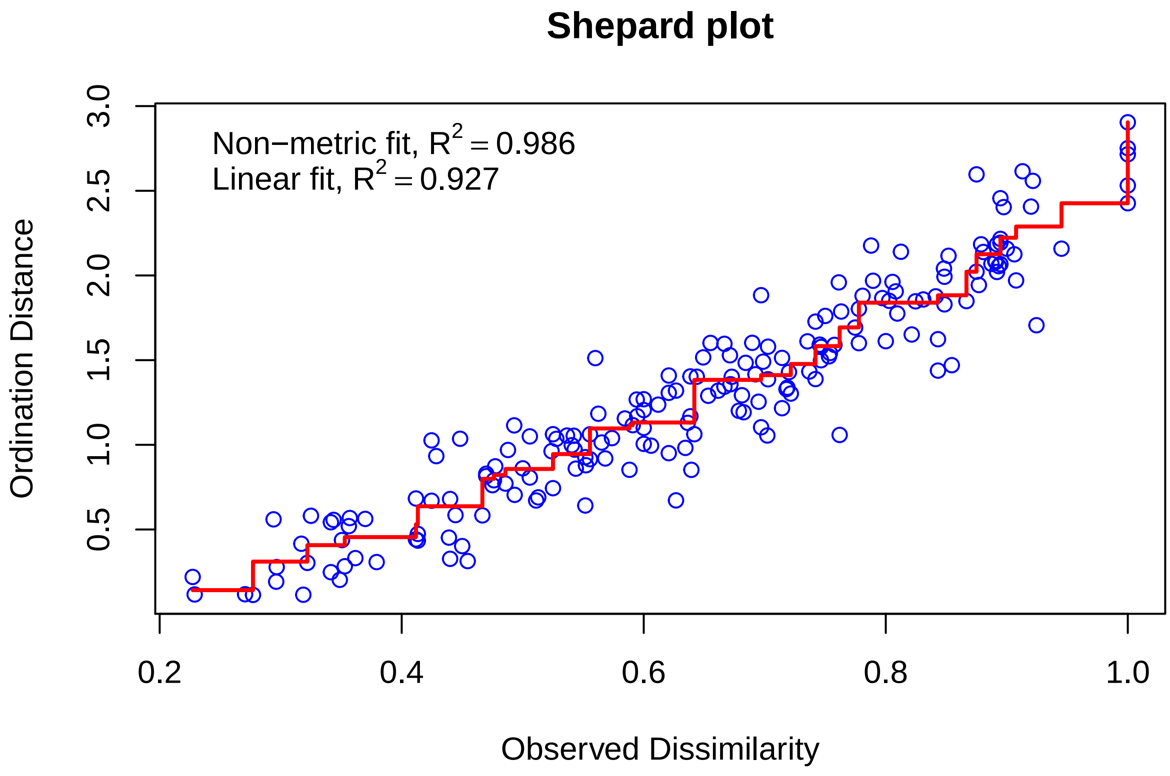
```
## Run 0 stress 0.1192678
## Run 1 stress 0.1183186
## ... New best solution
## ... procrustes: rmse 0.0202689  max resid 0.06495309
## Run 2 stress 0.1192682
## Run 3 stress 0.1192678
## Run 4 stress 0.1192679
## Run 5 stress 0.1183187
```

```
## ... procrustes: rmse 6.304471e-05  max resid 0.0001770544
## *** Solution reached
```

```r
dune_mds
```

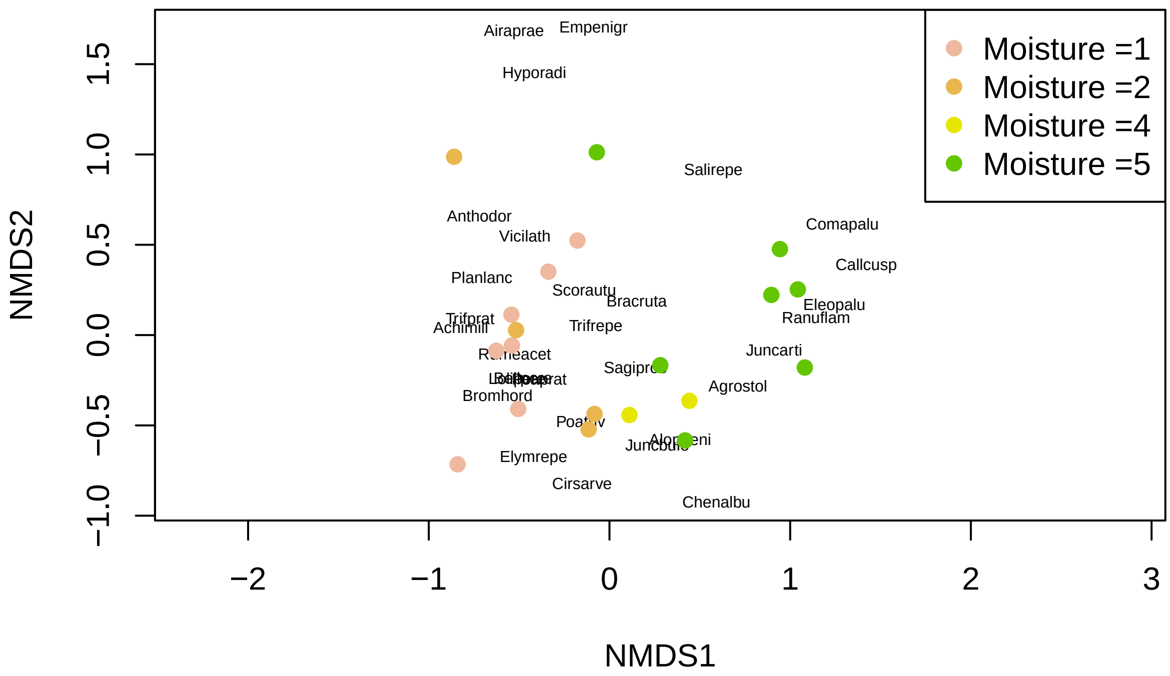
```
##
## Call:
## metaMDS(comm = dune)
##
## global Multidimensional Scaling using monoMDS
##
## Data:     dune
## Distance: bray
##
## Dimensions: 2
## Stress:     0.1183186
## Stress type 1, weak ties
## Two convergent solutions found after 5 tries
## Scaling: centring, PC rotation, halfchange scaling
## Species: expanded scores based on 'dune'
```

```r
stressplot(dune_mds, main="Shepard plot")
```




```r
plot(dune_mds, type='n')
text(dune_mds, 'sp', cex=.5)
# generate vector of colors
color_vect = rev(terrain.colors(6))[-1]
```

```r
points(dune_mds, 'sites', pch=19,
       col=color_vect[dune.env$Moisture])
legend('topright', paste("Moisture =", c(1,2,4,5), sep=''), #fixed
       col=color_vect, pch=19)
```



2. Continuing off the NMDS we can use direct ordination (CCA) to compare species assemblage and dune environmental data. I am interested in testing the hypothesis that dune plant communities differ significantly based on the availability of moisure. First, we construct a CCA model comparing dune species, and environmental data. CCA indicates that environmental conditions explain 71% of the variation in plant species assemblages. Graphically, moisture loads along both axes, potentially indicating a split (high vs low) in species distribution. An ANOVA indicates that this model fits significantly better than random observations (ie, observations are not likely to be random), but does not indicate significant partial effects of the environmental variables.

```r
#sorry for printing this!
r2_adj_cca = function(cca_obj, nperm, analytical=FALSE) {
      r2 = cca_obj$CCA$tot.chi / cca_obj$tot.chi
    n = nrow(cca_obj$CCA$Xbar)
    if (missing(nperm)) {
        # eq 4 Peres-Neto
        p = cca_obj$CCA$rank
        out = c(r2, 1 - (((n - 1) / (n - p - 1)) * (1 - r2)))
    }
    else {
        if (nperm <= 0)
            stop('nperm argument must either be a positive integer or not specified')
        if (!any(grepl('~', cca_obj$call)))
            stop('The model object must be specified using a model formula rather than providing indivi
        rand.r2 = rep(NA, nperm)
        Y_string = as.character(cca_obj$terms[[2]])
```

```
        Y = eval(parse(text=Y_string))
        for (i in 1:nperm) {
            Yrand = Y[sample(n), ]
            #assign(Y_string, Yrand)
            #cca.rand = eval(cca_obj$call)
            cca_obj$call[2] = sub('comm', 'Yrand', cca_obj$call[2])
            cca.rand = eval(parse(text=paste(cca_obj$call[1], '(',cca_obj$call[2],
                                            ', data=', cca_obj$call[3], ')',
                                            sep='')))
            #cca.rand = update(cca_obj, Yrand ~ .)
            rand.r2[i] = cca.rand$CCA$tot.chi / cca_obj$tot.chi
            if (i %% 100 == 0)  print(i)
        }
        # Eq 5 Peres-Neto
        out = c(r2,
                1 - ((1 - r2) / (1 - mean(rand.r2))))
    }
    names(out) = c('r2', 'r2adj')
    return(out)
}
```

```
dune_cca = cca(dune ~ . , data=dune.env) #compares the community matrix to the environmental matrix.
dune_cca #environmental variables explain more than 71% of variation
```

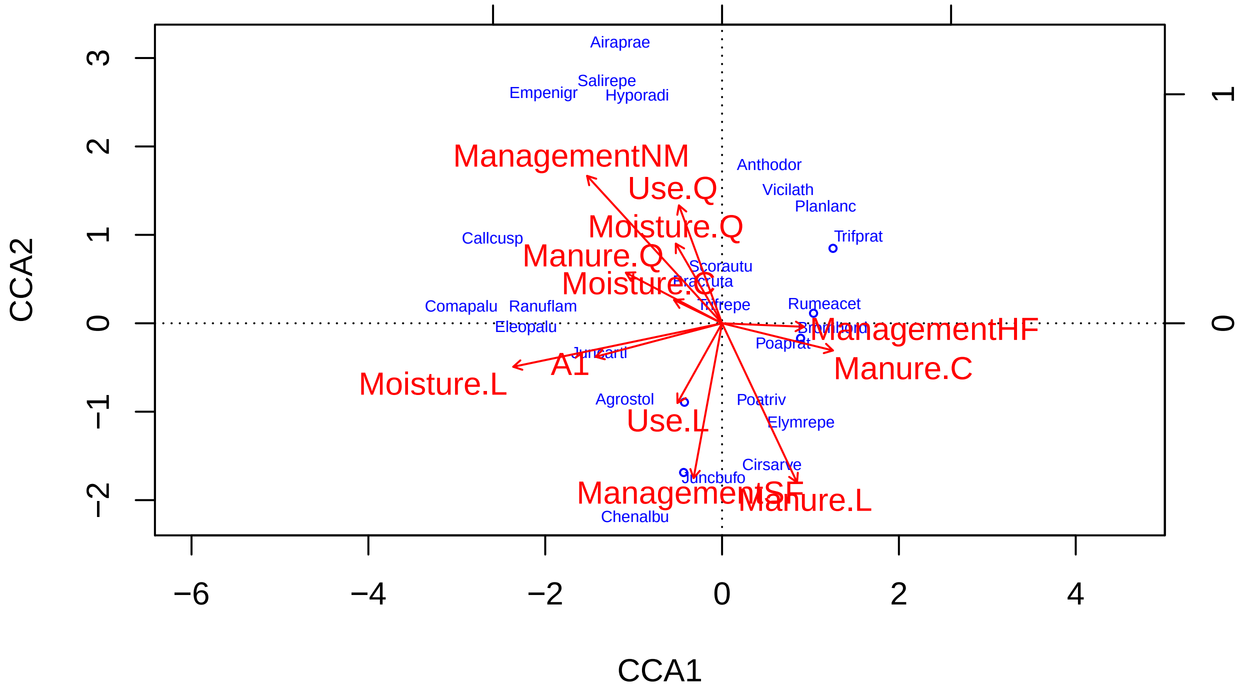
```
## Call: cca(formula = dune ~ A1 + Moisture + Management + Use +
## Manure, data = dune.env)
##
##                 Inertia Proportion Rank
## Total            2.1153     1.0000
## Constrained      1.5032     0.7106   12
## Unconstrained    0.6121     0.2894    7
## Inertia is mean squared contingency coefficient
## Some constraints were aliased because they were collinear (redundant)
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3   CCA4   CCA5   CCA6   CCA7   CCA8   CCA9  CCA10
## 0.4671 0.3410 0.1761 0.1532 0.0953 0.0703 0.0589 0.0499 0.0318 0.0260
##  CCA11  CCA12
## 0.0228 0.0108
##
## Eigenvalues for unconstrained axes:
##     CA1     CA2     CA3     CA4     CA5     CA6     CA7
## 0.27237 0.10876 0.08975 0.06305 0.03489 0.02529 0.01798
```

```
r2_adj_cca(dune_cca, 100, analytical = FALSE) # r2=.71
```

```
## [1] 100
```

```
##        r2     r2adj
## 0.7106267 0.0000000
```

```r
plot(dune_cca, type='n', scaling=1)
orditorp(dune_cca, display='sp', cex=0.5, scaling=1, col='blue')
text(dune_cca, display='bp', col='red')
```




```r
#moisture is somewhat split in loading along both axies. could this support my high vs low moisture hyp

anova(dune_cca)
```

```
## Permutation test for cca under reduced model
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ A1 + Moisture + Management + Use + Manure, data = dune.env)
##          Df ChiSquare      F Pr(>F)
## Model    12    1.5032 1.4325  0.022 *
## Residual  7    0.6121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(dune_cca, by='margin')
```

```
## Permutation test for cca under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ A1 + Moisture + Management + Use + Manure, data = dune.env)
##            Df ChiSquare      F Pr(>F)
## A1          1   0.11070 1.2660  0.266
```

```
## Moisture    3   0.31587 1.2041  0.194
## Management  2   0.15882 0.9081  0.561
## Use         2   0.13010 0.7439  0.787
## Manure      3   0.25490 0.9717  0.484
## Residual    7   0.61210
```

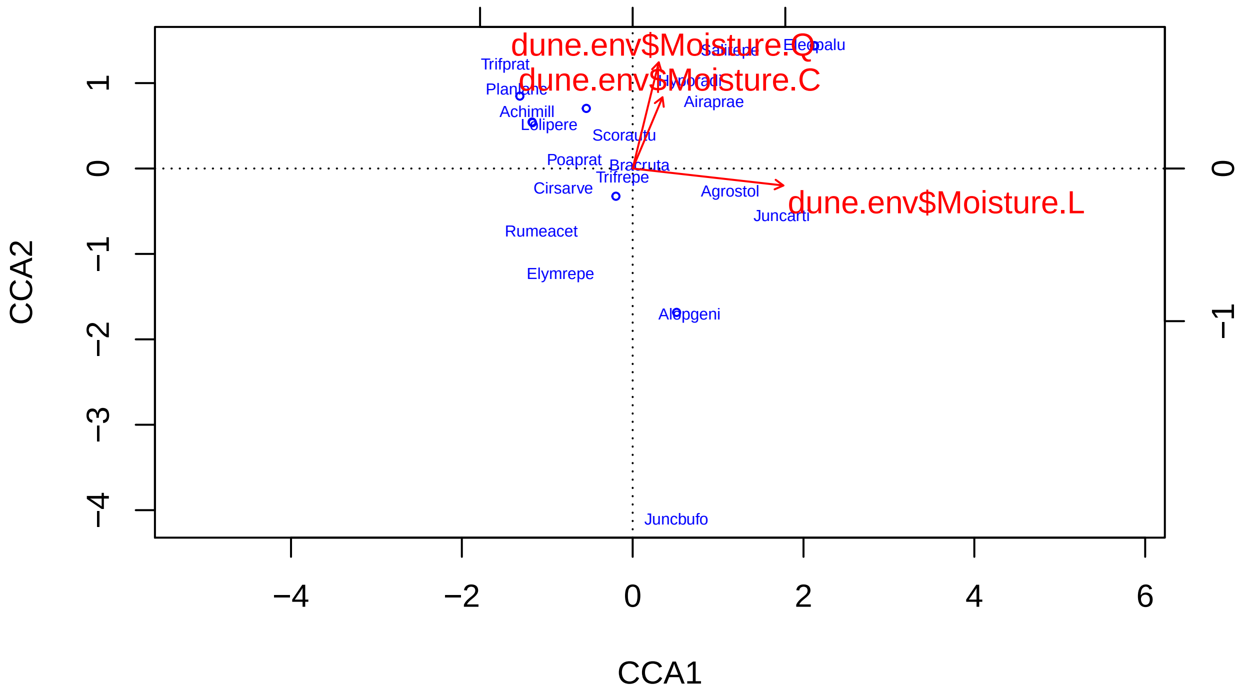
```
#the model itself is significant
```

Because I am most interested in the role of moisture as an explanatory variable, an addational CCA using only moisture was conducted. This model indicates that moisture alone explains 30% of the variation explained by the environmental variables, or 40% (.297/.7106) of the variation in plant species composition. Plotting this model again indicates a split in how moisture loads on the axes, and an ANOVA indicates a significant difference in plant species composition between moisture levels.

```
dune_cca_moist = cca(dune ~ dune.env$Moisture)
dune_cca_moist #moisture alone explains 30% of variation or.
```

```
## Call: cca(formula = dune ~ dune.env$Moisture)
##
##                Inertia Proportion Rank
## Total           2.1153     1.0000
## Constrained     0.6283     0.2970    3
## Unconstrained  1.4870     0.7030   16
## Inertia is mean squared contingency coefficient
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3
## 0.4187 0.1330 0.0766
##
## Eigenvalues for unconstrained axes:
##    CA1    CA2    CA3    CA4    CA5    CA6    CA7    CA8    CA9   CA10
## 0.4098 0.2259 0.1761 0.1234 0.1082 0.0908 0.0859 0.0609 0.0566 0.0467
##   CA11   CA12   CA13   CA14   CA15   CA16
## 0.0419 0.0201 0.0143 0.0099 0.0085 0.0080
```

```
plot(dune_cca_moist, type='n', scaling=1)
orditorp(dune_cca_moist, display='sp', cex=0.5, scaling=1, col='blue')
text(dune_cca_moist, display='bp', col='red')
```




```r
anova(dune_cca_moist)
```

```
## Permutation test for cca under reduced model
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ dune.env$Moisture)
##          Df ChiSquare      F Pr(>F)
## Model     3   0.62831 2.2536  0.003 **
## Residual 16   1.48695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. It seems that the indirect and direct ordination methods are telling the same story but in different ways. The take home message is the same, plant species composition can be explained by moisture, but indirect ordination shows this visually and is useful for exploration of trends in the data. Direct ordination is able to show this same trend numerically and with stastical tests, but is only useful after general exploration of the data has been conducted. Ultimately, direct ordination is more useful, but requires the explorative tools of indirect methods.