Stat 462 Lab 9 solutions

March 18, 2014

Question 12.1

First, divide the possum data into females and males

```
possum.female <- possum[ possum$sex == "f", ]
possum.male <- possum[ possum$sex == "m", ]
```

Now run principal components on each sex and both together

```
possum.f.prc <- princomp( na.omit( possum.female[,6:14] ) )
possum.m.prc <- princomp( na.omit( possum.male[,6:14] ) )
possum.prc <- princomp( na.omit( possum[,6:14] ) )
```

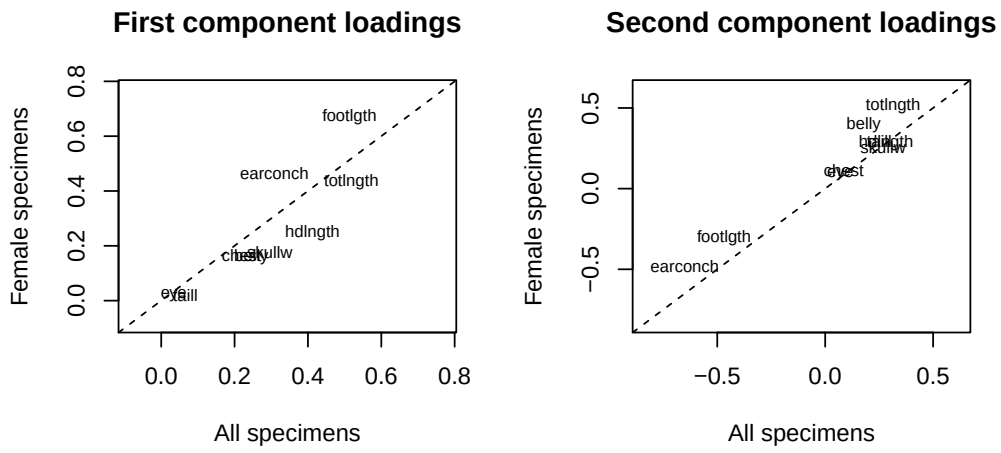
For each of the first and second principal components, plot the loadings for females against the loadings for all data combined, and similary for males. The principal components need to be *rotated* by flipping the signs of some of the loadings for the plot. Visual inspection reveals that both female loadings need to be flipped and the first male loading needs to be flipped to match the combined principal components loadings.

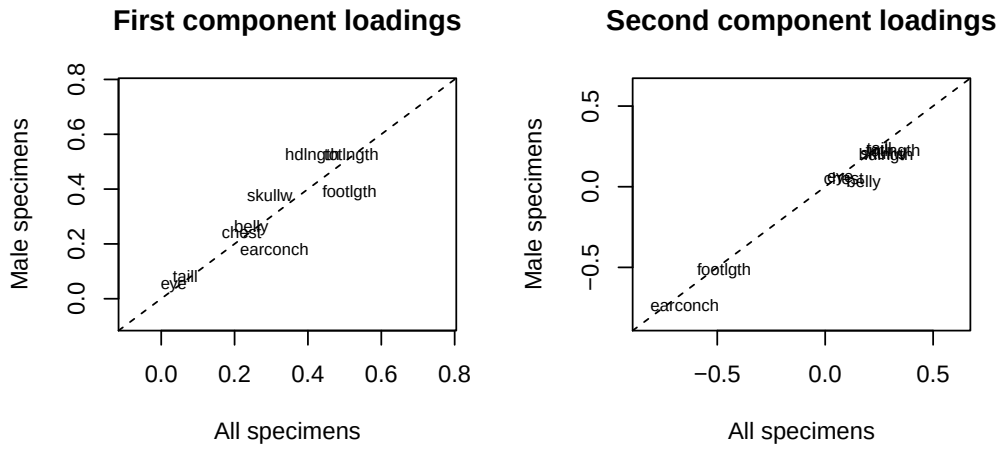
```
f.loadings <- possum.f.prc$loadings[,1:2]
m.loadings <- possum.m.prc$loadings[,1:2]
all.loadings <- possum.prc$loadings[,1:2]
# Flip some of the loadings...
f.loadings <- f.loadings*(-1)
m.loadings[,1] <- m.loadings[,1]*(-1)
```

Now we plot...

```
# Get the range of the loadings for the plots
lrange1 <- range( f.loadings[,1], m.loadings[,1], all.loadings[,1] )+c(-0.1,0.1)
lrange2 <- range( f.loadings[,2], m.loadings[,2], all.loadings[,2] )+c(-0.1,0.1)
par( mfrow = c(2,2) )
# Females first component
plot( all.loadings[,1], f.loadings[,1], xlab = "All specimens"
     , ylab = "Female specimens", main = "First component loadings"
     , xlim = lrange1, ylim = lrange1, pch = "" )
text( all.loadings[,1], f.loadings[,1], rownames(f.loadings), cex = 0.7 )
abline( a = 0, b = 1, lty = "dashed" )

# Females second component
plot( all.loadings[,2], f.loadings[,2], xlab = "All specimens"
     , ylab = "Female specimens", main = "Second component loadings"
     , xlim = lrange2, ylim = lrange2, pch = "" )
```

```
text( all.loadings[,2], f.loadings[,2], rownames(f.loadings), cex = 0.7 )
abline( a = 0, b = 1, lty = "dashed" )

# Males first component
plot( all.loadings[,1], m.loadings[,1], xlab = "All specimens"
    , ylab = "Male specimens", main = "First component loadings"
    , xlim = lrange1, ylim = lrange1, pch = "" )
text( all.loadings[,1], m.loadings[,1], rownames(m.loadings), cex = 0.7 )
abline( a = 0, b = 1, lty = "dashed" )

# Males second component
plot( all.loadings[,2], m.loadings[,2], xlab = "All specimens"
    , ylab = "Male specimens", main = "Second component loadings"
    , xlim = lrange2, ylim = lrange2, pch = "" )
text( all.loadings[,2], m.loadings[,2], rownames(m.loadings), cex = 0.7 )
abline( a = 0, b = 1, lty = "dashed" )
```





Interpretation of these plots can be tricky. It looks like the females differ from the combined principal components. In particular, in the first component the foot length and ear conch loadings are larger the loadings for all specimens, and the head length and skull width loadings are smaller than the loadings for all specimens. Since the first component is roughly a metric of overall size, we might interpret this as meaning that there is more variation in ear conch and footlength for females compared to all specimens, and less variation in head length and skull width for females compared to all specimens. In the second component, the ear conch and foot length loadings are not as large (less negative) for females compared to all specimens and the total length and belly loadings are larger for females compared to all specimens. The second component is roughly a metric of the difference between ear conch and foot length measurements compared to other measurements. The smaller ear conch and foor length loadings for females in the second component might be offsetting their larger importance in describing the overall size (i.e. the first component).

Question 12.4

The painters data.frame has scores for composition, drawing, colour, and expression.

The first step is to calculate a distance (or *dissimilarity*) matrix for the painters.

```
library(MASS)
paint.dist <- dist( painters[,-5] )
```

From the dissimilarity matrix, we use classical metric scaling to get a two-dimensional representation of the dissimilarities between the painters. Classical metric scaling tries to preserve distances as much as possible when computing the 2d representation.

```
paint.cmd <- cmdscale( paint.dist, k = 2 )
```

If we try to use Sammon scaling (which does a weighted version of classical MDS – weights proportional to the dissimilarities so that large dissimilarities count more than small dissimilarities in determining the 2d representation), we get an error because two of the artists have identical scores (zero dissimilarity), causing an error in the weights. You might drop one of the offending aritists, or perhaps fudge their score a bit to create a small dissimilarity...

```
painters2 <- painters
painters2[2,1] <- painters2[2,1] + 1
paint.dist2 <- dist( painters2[,-5] )
paint.sam <- sammon( paint.dist2 )
```

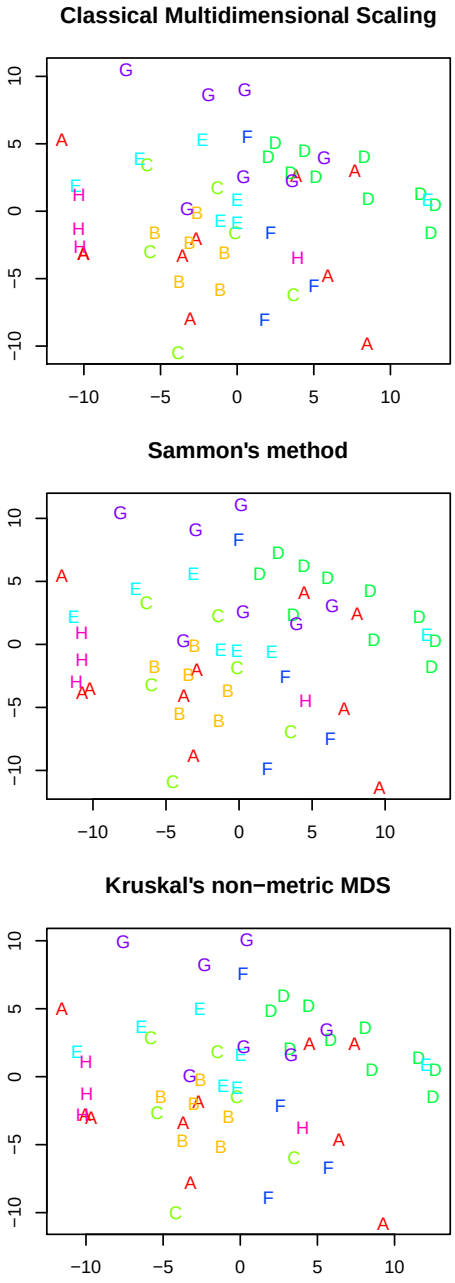
Finally, Kruskal's non-metric scaling (which also requires non-zero dissimilarities)

```
paint.iso <- isoMDS( paint.dist2, k = 2 )
```

Now plot the 2d representations for each scaling method

```
par(mfrow = c(3,1), mar = c(2,3,4,1) )

plot( paint.cmd, xlab = "", ylab = ""
    , pch = as.character(painters[,5])
    , col = rainbow(8)[as.integer(painters[,5])]
    , main = "Classical Multidimensional Scaling" )

plot( paint.sam$points, xlab = "", ylab = ""
    , pch = as.character(painters[,5])
    , col = rainbow(8)[as.integer(painters[,5])]
    , main = "Sammon's method" )
```

```
plot( paint.iso$points, xlab = "", ylab = ""
    , pch = as.character(painters[,5])
    , col = rainbow(8)[as.integer(painters[,5])]
    , main = "Kruskal's non-metric MDS" )
```



12.4 additional distance measures (862 students)

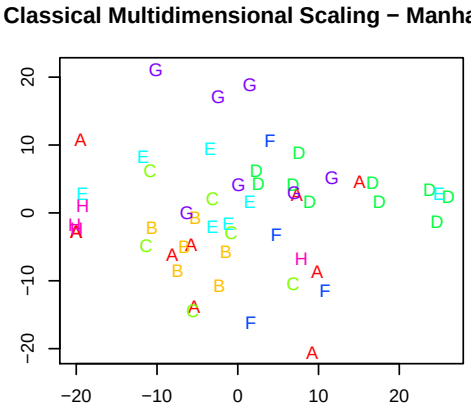
The analysis proceeds exactly the same as above, except use Manhattan

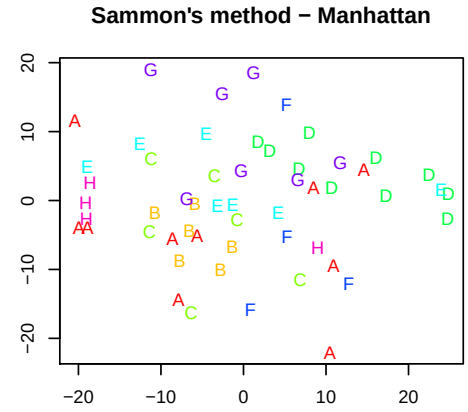
```
paint.dist <- dist( painters[,-5], method = "manhattan" )
```

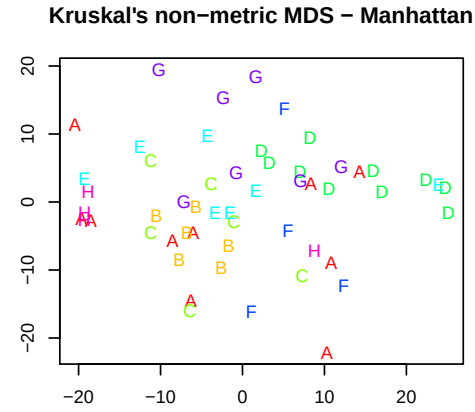
or canberra distances

```
paint.dist <- dist( painters[,-5], method = "canberra" )
```

The scaling based on Manhattan distances are very similar to Euclidean distances







The scaling based on the canberra distances are noticeably different, however.