phyloExample

2018-02-06

Load the packages

```r
library(Rcompadre)
library(ape)
library(caper)
library(popbio)
```

We will make use of the mammalian supertree.

```r
mamtree <- read.nexus("/Users/jones/Dropbox/ESA\ 2017\ Portland\ Workshop/data/mammaltree.nex")
summary(mamtree)
```

```
>
> Phylogenetic tree: mamtree
>
>   Number of tips: 5139
>   Number of nodes: 2458
>   Branch lengths:
>     mean: 8.852233
>     variance: 67.23214
>     distribution summary:
>     Min.  1st Qu.   Median  3rd Qu.     Max.
>   0.0089   3.4000   6.7000  11.9000 102.6000
>   No root edge.
>   First ten tip labels: Tachyglossus_aculeatus
>                         Zaglossus_bruijni
>                         Zaglossus_attenboroughi
>                         Zaglossus_bartoni
>                         Ornithorhynchus_anatinus
>                         Anomalurus_beecrofti
>                         Anomalurus_derbianus
>                         Anomalurus_pelii
>                         Anomalurus_pusillus
>                         Zenkerella_insignis
>   No node labels.
```

Subsetting the data

Firstly, subset to mean matrices, for mammals, that have a duration of 3+ and dimension of >3.

```r
x<-Rcompadre::subsetDB(comadre, MatrixComposite == "Mean" &
                        Class == "Mammalia" &
                        StudyDuration >= 3 &
                        MatrixDimension > 3)
```

The straightforward phylogenetic regression models require one measure per species. So I will take only the first entry for each species if there are duplicates. In addition, I want to subset to only those species that are actually in our phylogenetic tree.

```r
temp <- which(!duplicated(x$metadata$SpeciesAccepted))
rownames(x$metadata)<-NULL
y<-Rcompadre::subsetDB(db=x,sub = rownames(x$metadata) %in% temp &
            SpeciesAccepted %in% mamtree$tip.label)
```

So now we have a version of COMADRE called `y` that contains only the species of interest.

Now we can loop through those species and make some calculations. Here I am calculating life expectancy from the start of the first stage, which is the sum of the first column of the fundamental matrix.

This code looks a little complicated because it is not possible to calculate the fundamental matrix for some matrices. Therefore I have to add an `if` statement to cope with those problems.

```
demog <- data.frame(species = y$metadata$SpeciesAccepted)
demog$lexp <- NULL
for(i in 1:nrow(y$metadata)){
  A <- y$mat[[i]]$matA
  fm <- fundamental.matrix(A)
  if(class(fm) == "list"){
  demog$lexp[i]<-sum(fm$N[,1])}else{demog$lexp[i] <- NA}
}
```

Now I have a data frame called `demog` that includes the species name and life expectancy.

Now I can import weight data for these species.

```
mammweight<-read.csv("/Users/jones/Dropbox/ESA\ 2017\ Portland\ Workshop/data/mammalweight.csv")
```

And merge it with the existing data.

```
demog2 <- merge(demog,mammweight,by="species")
demog2$logmass <- log(demog2$weight)
demog2$loglexp <- log(demog2$lexp)
```

I can now plot the relationship between log mass and log life expectancy.

```
plot(demog2$logmass,demog2$loglexp)
```
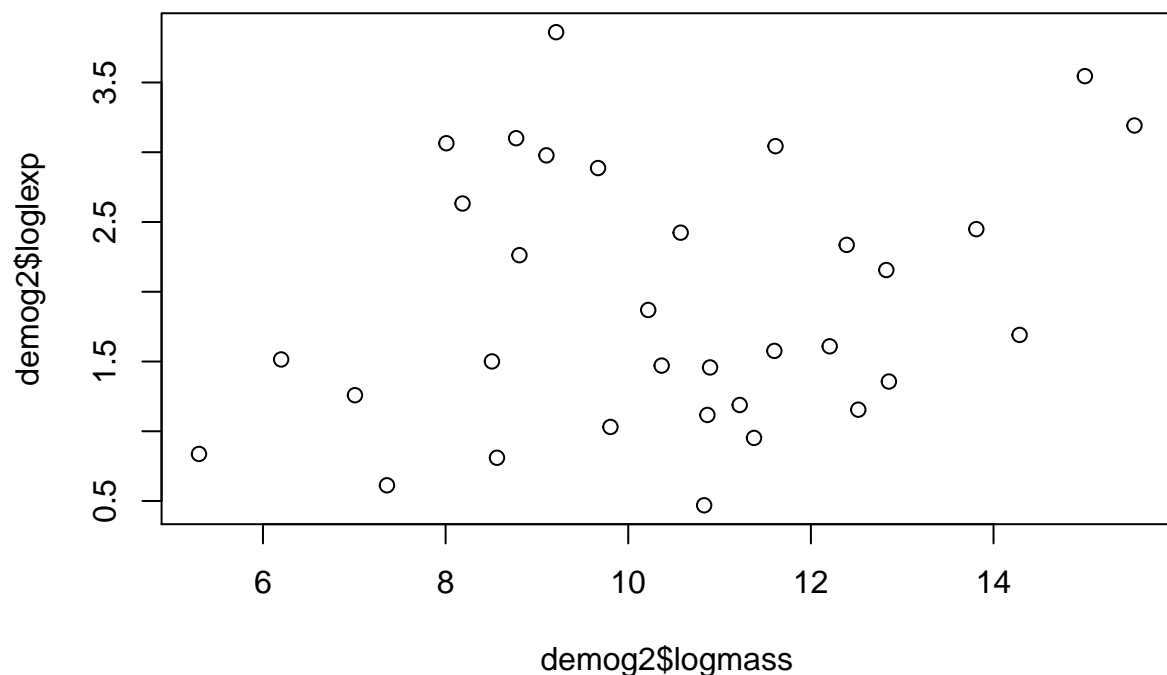

For the next bit we will use functions from `caper`. We make a "comparative data object"

```
mamdat <- caper::comparative.data(data = demog2, phy = mamtree, names.col = "species",na.omit=FALSE)
```

Fit a PGLS model of the relationship between longevity and weight.

```
mod1<-caper::pgls(loglexp~logmass,mamdat,lambda="ML")
summary(mod1)
```

```
>
```

```
> Call:
> caper::pgls(formula = loglexp ~ logmass, data = mamdat, lambda = "ML")
>
> Residuals:
>      Min       1Q   Median       3Q      Max
> -0.14402 -0.05059  0.03478  0.06212  0.15284
>
> Branch length transformations:
>
> kappa  [Fix]  : 1.000
> lambda [ ML]  : 0.750
>     lower bound : 0.000, p = 0.00062576
>     upper bound : 1.000, p = 0.00061216
>     95.0% CI    : (0.369, 0.942)
> delta  [Fix]  : 1.000
>
> Coefficients:
>               Estimate Std. Error t value Pr(>|t|)
> (Intercept) -0.104460   0.802064 -0.1302 0.897219
> logmass      0.198547   0.068865  2.8831 0.007095 **
> ---
> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> Residual standard error: 0.07945 on 31 degrees of freedom
>   (2 observations deleted due to missingness)
> Multiple R-squared: 0.2114,   Adjusted R-squared: 0.186
> F-statistic: 8.312 on 1 and 31 DF,  p-value: 0.007095
```

Let's have a look at the likelihood profile.

```
plot(caper::pgls.profile(mod1))
```


Data: mamdat; Model: loglexp ~ logmass
kappa 1.00; lambda 0.75; delta 1.00
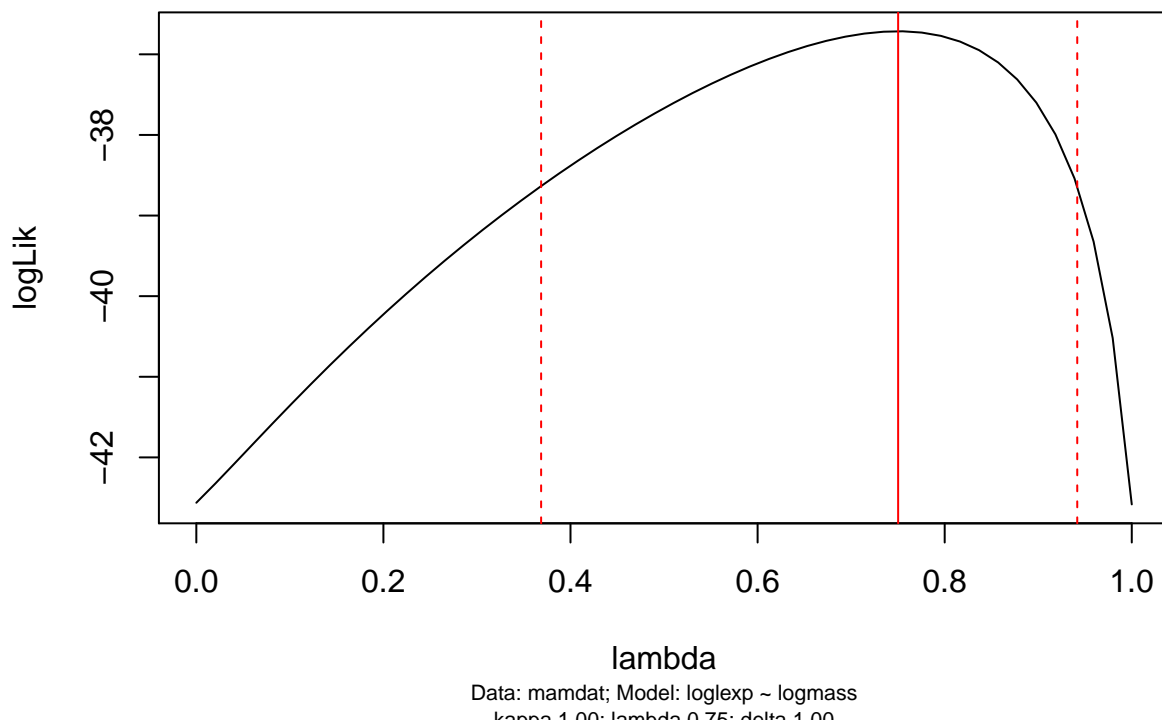
Let's plot this regression, and compare it to an ordinary least squares regression.

```
plot(demog2$logmass,demog2$loglexp)
abline(mod1)
```

```
modols<-lm(loglexp~logmass,mamdat$data)
abline(modols,col="red")
```