Daniella Assing

Geosci 541

March 14 2016

Lab #7

Calculating stratigraphic ranges

#### **Problem Set 1**

There are four columns in DataPBDB relevant to the age of an organism: early\_interval, late\_interval, max\_ma, and min\_ma. Because we rarely have a precise date, we generally give the age of an occurrence as a range. This range can be expressed by interval names or by numbers.

1) What do the max\_ma and min\_ma columns of DataPBDB represent? If you do not intuitively know, you can always check the [Paleobiology Database API documentation](https://paleobiodb.org/data1.2/occs/list_doc.html).

The max\_ma column in this dataset represents the maximum age in Ma and the min\_ma column represents the minimum age of the sample in Ma (i.e. the stratigraphic range).

2) What is oldest age of each genus? [[Hint](https://github.com/aazaff/startLearn.R/blob/master/intermediateConcepts.md#direct-subsetting-with-functionals): Use the tapply( ) and max( ) functions we've used in previous labs]. Show the code you would use to find out.

>tapply(DataPBDB[,"max\_ma"],DataPBDB[,"genus"],max)

3) What is the youngest age of each genus? [[Hint](https://github.com/aazaff/startLearn.R/blob/master/intermediateConcepts.md#direct-subsetting-with-functionals): Use the tapply( ) and max( ) functions we've used in previous labs]. Show the code you would use to find out.

>tapply(DataPBDB[,"min\_ma"],DataPBDB[,"genus"],min)

4) Find which genus has the most occurrences in the dataset [Hint: Use the table( ) function!]. What code did you use?

>sort(table(DataPBDB[,"genus"]))

Anadara

1916

5) What is the stratigraphic range of this taxon (i.e., your answer to question 4). Show your code.

> Anadara<-DataPBDB[which(DataPBDB[,26] =="Anadara"),]

> max(Anadara[,15])

[1] 66

> min(Anadara[,16])

[1] 0

> max(Anadara[,15])-min(Anadara[,16])

[1] 66

#### **Problem Set 2**

1) Qualitatively describe what is happening in the following line of code. A good answer should identify what the different arguments are for each function, and what they are used for.

mean(sample(Lucina[,"paleolng"],length(Lucina[,"paleolng"]),replace=TRUE))

The first argument finds the mean of the data set. The second argument tells R to take a sample of a specific size of the paleolocation of Lucina occurrences. The third argument tells R what size to include of the paleolocations of Lucina occurrences and the final argument replace=TRUE allows for elements to be resampled again from the same dataset, that is, if an element has already been picked randomly it can be picked again in a new trial.

2) Plot a [kernel density](https://github.com/aazaff/startLearn.R/blob/master/expertConcepts.md#describing-distributions-with-statistics) graph of ResampledMeans. Show your code. Does the distribution look approximately Gaussian? Explain why you think it does or does not.

> plot(density(ResampledMeans))

I think it looks Gaussian because it seems centered around a particular mean, however it is not perfectly Gaussian because there is a little asymmetry.

3) Find the mean of ResampledMeans, is it similar to the mean of the original data?

> mean(ResampledMeans)

[1] 24.16227

Yes it’s very similar to the mean of the original data.

4) Sort ResampledMeans from lowest to highest. [Hint: We learned how to sort a vector in [Lab 6](https://github.com/aazaff/teachPaleobiology/blob/master/LabExercise6.md#problem-set-2)].

>sort(ResampledMeans)

5) Now that you have sorted ResampledMeans, what is the 2.5th percentile of ResampledMeans and what is the 97.5th percentile of Resampled means. If you do not know what a percentile is, or how to calculate it, you can use google. Show your code.

> quantile(ResampledMeans, c(.025, .975))

2.5% 97.5%

21.82094 26.28630

6) Incidentally, these numbers (your answer to question 5) are the lower and upper confidence interval of the mean! Qualitatively explain why this is the case.

This means that there should 95% certainty that the mean of the dataset lies between these two intervals (which it does).

#### **Problem Set 3**

1) Based on the confidence intervals given above, do you think it likely or unlikely that *Lucina* is still alive?

I think it is likely that Lucina is still alive because the confidence intervals for the extinction dates are calculated. The earliest extinction date is 0.00 which most likely means that it is not yet extinct. I also assume that the negative number for the latest extinction means the future date at which it is predicted to be extinct.

2) Find the extinction confidence interval for the genus *Dallarca*. Show your code.

> Dallarca<-subset(DataPBDB,DataPBDB[,"genus"]=="Dallarca")

> OriginalMeanDallarca<-mean(Dallarca[,"paleolat"])

> OriginalMeanDallarca

[1] 38.84449

> set.seed(100)

>NewMeanDallarca<-mean(sample(Dallarca[,"paleolat"],length(Dallarca[,"paleolat"]),replace=TRUE))

> NewMeanDallarca

[1] 39.20812

^^^^^^^ Maybe this part above wasn’t necessary? Comments?

> estimateExtinction <- function(OccurrenceAges, ConfidenceLevel=.95) {

+ # Find the number of unique "Horizons"

+ NumOccurrences<-length(unique(OccurrenceAges))-1

+ Alpha<-((1-ConfidenceLevel)^(-1/NumOccurrences))-1

+ Lower<-min(OccurrenceAges)

+ Upper<-min(OccurrenceAges)-(Alpha\*10)

+ return(setNames(c(Lower,Upper),c("Earliest","Latest")))

+ }

> estimateExtinction(Dallarca[,"min\_ma"],0.95)

Earliest Latest

2.58800 -3.88749

3) A pure reading of the fossil record says that *Dallarca* went extinct at the end of the Pliocene Epoch. Based on its confidence interval, do you think it is possible that *Dallarca* is still extant (alive)?

Based only on the confidence interval I would say yes, it’s possible that it is still extant (negative CI for latest).

4) In this case, should we trust the confidence interval or a pure reading of the fossil record? Explain your reasoning.

Maybe neither? Or be cautious of them both equally. The confidence interval assumes random distribution in space and time which isn’t true for many organisms because they may prefer specific environments so this isn’t random distribution. A pure reading of the fossil record can help but should also not be trusted because there have been cases where fossils have reappeared after being assumed to have gone extinct. This can be seen with Lazarus and Zombie taxa (not sure where I’m going with this!). I also think of the coelacanth as another reason not to completely trust the pure record because it was thought to be extinct when in fact it was not.

#### **Problem Set 4**

1) State one ecological reason why this assumption is unlikely to be true.

Organisms may prefer certain environments so they are definitely not randomly distributed in time and space. A good example of this is the benthic community that lives in methane seeps.

2) State one geological reason why this assumption is unlikely to be true.

A population may be live in a habitat with geologic barriers and this would concentrate them in one location. One example I can think of is Jellyfish Lake on Eil Malk island in Palau. The lake is geologically isolated with few openings and the jellyfish here form their own species.

#### 

#### **Problem Set 5**

1) How many occurrences are in DataPBDB. How many are in ExtantData? How many occurrences were lost by limiting our analysis to only extant bivalves?

> dim(DataPBDB)

[1] 67618 26

> dim(ExtantData)

[1] 59097 26

> dim(DataPBDB)-dim(ExtantData)

[1] 8521 0

2) How many unique( ) genera were in DataPBDB and ExtantData, respectively. Using this information, what percentage of Cenozoic bivalves in the PBDB are still extant today.

> sum(DataPBDB[,"genus"] != 0)

[1] 67618

> sum(ExtantData[,"genus"] != 0)

[1] 59097

> sum(DataPBDB[,"genus"] != 0)-sum(ExtantData[,"genus"] != 0)

[1] 8521

> 67618-59097

[1] 8521

> 8521/67618

[1] 0.1260167

> 0.1260167\*100

[1] 12.60167

= 12.6%

3) Find the stratigraphic range of fossil occurrences for each genus in the ExtantData dataset. If you do not remember how to do this, revisit [Problem Set 1](https://github.com/aazaff/teachPaleobiology/blob/master/LabExercise7.md#problem-set-1) of this lab.

>tapply(ExtantData[,"max\_ma"],ExtantData[,"genus"],max)-tapply(ExtantData[,"min\_ma"],ExtantData[,"genus"],min)

4) Using your answer to question 3, find which genera in ExtantData are not extant according to the PBDB - i.e., do not have a minimum min\_age of zero. Show your code.

> ExtantMin<-tapply(ExtantData[,"min\_ma"],ExtantData[,"genus"],min)

>which((ExtantMin) > 0)

> NotExtant<-which((ExtantMin) > 0)

> length(NotExtant)

[1] 306

5) Calculate stratigraphic confidence intervals for the following genera (careful with your spelling!): *Scorbicularia*, *Meiocardia*,*Dimya*, *Digitaria*, *Cuspidaria*, *Arctica*, *Aloides*, *Kurtiella*, *Gouldia*, and *Acrosterigma*. Show your code. What percentage of these taxa have confidence intervals indicating that the taxon might still be extant?

> Scorbicularia<-subset(DataPBDB,DataPBDB[,"genus"]=="Scorbicularia")

> Meiocardia<-subset(DataPBDB,DataPBDB[,"genus"]=="Meiocardia")

> Dimya<-subset(DataPBDB,DataPBDB[,"genus"]=="Dimya")

> Digitaria<-subset(DataPBDB,DataPBDB[,"genus"]=="Digitaria")

> Cuspidaria<-subset(DataPBDB,DataPBDB[,"genus"]=="Cuspidaria")

> Arctica<-subset(DataPBDB,DataPBDB[,"genus"]=="Arctica")

> Aloides<-subset(DataPBDB,DataPBDB[,"genus"]=="Aloides")

> Kurtiella<-subset(DataPBDB,DataPBDB[,"genus"]=="Kurtiella")

> Gouldia<-subset(DataPBDB,DataPBDB[,"genus"]=="Gouldia")

> Acrosterigma<-subset(DataPBDB,DataPBDB[,"genus"]=="Acrosterigma")

> estimateExtinction <- function(OccurrenceAges, ConfidenceLevel=.95) {

+ # Find the number of unique "Horizons"

+ NumOccurrences<-length(unique(OccurrenceAges))-1

+ Alpha<-((1-ConfidenceLevel)^(-1/NumOccurrences))-1

+ Lower<-min(OccurrenceAges)

+ Upper<-min(OccurrenceAges)-(Alpha\*10)

+ return(setNames(c(Lower,Upper),c("Earliest","Latest")))

+ }

> estimateExtinction(Scorbicularia[,"min\_ma"],0.95)

Earliest Latest

Inf Inf

> estimateExtinction(Meiocardia[,"min\_ma"],0.95)

Earliest Latest

0.011700 -5.329574

> estimateExtinction(Dimya[,"min\_ma"],0.95)

Earliest Latest

0.781000 -2.054688

> estimateExtinction(Digitaria[,"min\_ma"],0.95)

Earliest Latest

0.781000 -3.761154

> estimateExtinction(Cuspidaria[,"min\_ma"],0.95)

Earliest Latest

2.5880000 0.7771965

> estimateExtinction(Arctica[,"min\_ma"],0.95)

Earliest Latest

0.011700 -1.799104

> estimateExtinction(Aloides[,"min\_ma"],0.95)

Earliest Latest

5.333 -Inf

> estimateExtinction(Kurtiella[,"min\_ma"],0.95)

Earliest Latest

0.01170 -34.70966

> estimateExtinction(Gouldia[,"min\_ma"],0.95)

Earliest Latest

0.011700 -2.047386

> estimateExtinction(Acrosterigma[,"min\_ma"],0.95)

Earliest Latest

0.011700 -3.481128

#Percent extant

> (8/10)\*100

[1] 80