Report

Keliang Xu, Shuting Li, Tao He

3/2/2022

1 Abstract

This report consists of 4 main sections: Introduction, Checking and Testing, Discussion, and Conclusion.

2 Introduction

2.1 Data Set Introduction

The variables of interest are "Taxa", "Site" and "Carbon and Nitrogen Level' '. The "Taxa ' ' includes four types of animals, which are turkey, cottontails, deer, and hare, respectively. The "site" includes two cities, which are BAOX and PATT.

The "Carbon and Nitrogen level" includes stable carbon and nitrogen isotopes from two closely related sites (BAOX and PATT) from the same area and period. These enter bones by eating and reflect the isotopic values of food sources. Carbon isotopes tell us about the level of human-grown plants in the diet, such as maize (corn). Carbon isotopes from collagen are protein-derived. Carbon isotopes from apatite (carbonate) reflect the whole diet. Nitrogen isotopes tell us about the protein in the diet.

2.2 Client's Work

So far, the client has:

1.) Use ANOVA to test for the differences between BAOX and PATT for each species.

2.) Examine box-plots and test just those groupings that look different.

2.3 Client Requests and Deliverables

1.) Check all her works, including the summary data table and box plots.

2.) Use statistical methods to show there is no significant difference in specific animals' diets from BAOX and PATT, except cottontails.

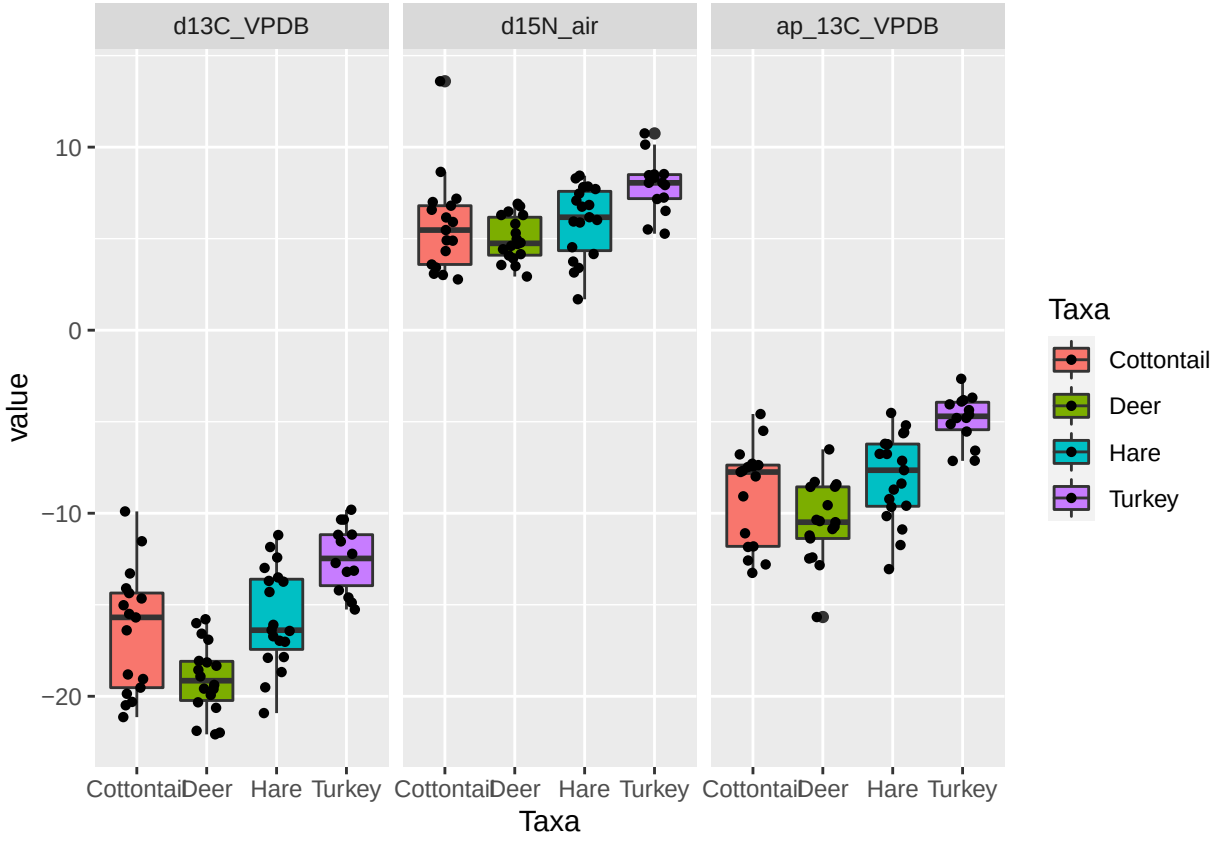
3 Checking and Testing

3.1 Checking

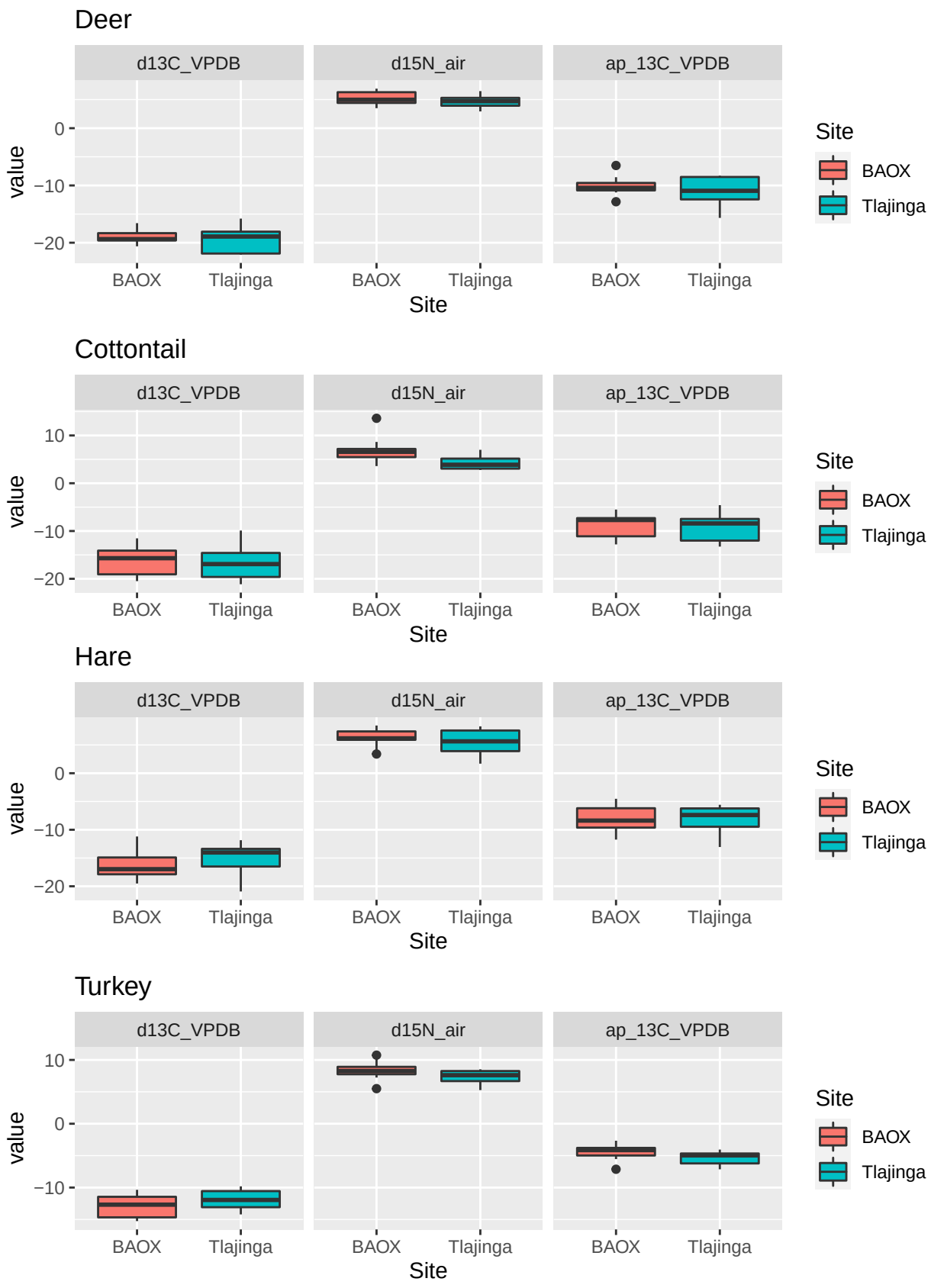
Dropped 4 missing value, their IDs are MC381,MC389,MC393,MC76.

EDA

In this plot, we compared specific isotopes level in 4 taxon. It showed the level of d15N_air is less diverse than other 2 isotopes.



Then we compared isotopes in two sites for each kind of animals. The plot shows that the median level of d13C_VPDB and d15N_air in BAOX is higher than Tlajinga of cottontail. Similarly, the median level of d13C_VPDB and d15N_air of hare also show difference in the boxplot.


Deer
d13C_VPDB
d15N_air
ap_13C_VPDB
value
0
−10
−20
BAOX
Tlajinga
Site
Site
BAOX
Tlajinga
Cottontail
d13C_VPDB
d15N_air
ap_13C_VPDB
value
10
0
−10
−20
BAOX
Tlajinga
Site
Site
BAOX
Tlajinga
Hare
d13C_VPDB
d15N_air
ap_13C_VPDB
value
0
−10
−20
BAOX
Tlajinga
Site
Site
BAOX
Tlajinga
Turkey
d13C_VPDB
d15N_air
ap_13C_VPDB
value
10
0
−10
BAOX
Tlajinga
Site
Site
BAOX
Tlajinga

Check Outlier

```
mean(filter(aggdata,Taxa=="Hare" & Site=="BAOX" & variable=="d15N_air")$value)
```

```
## [1] 6.276364
```

```
sd(filter(aggdata,Taxa=="Hare" & Site=="BAOX" & variable=="d15N_air")$value)
```

```
## [1] 1.582891
```

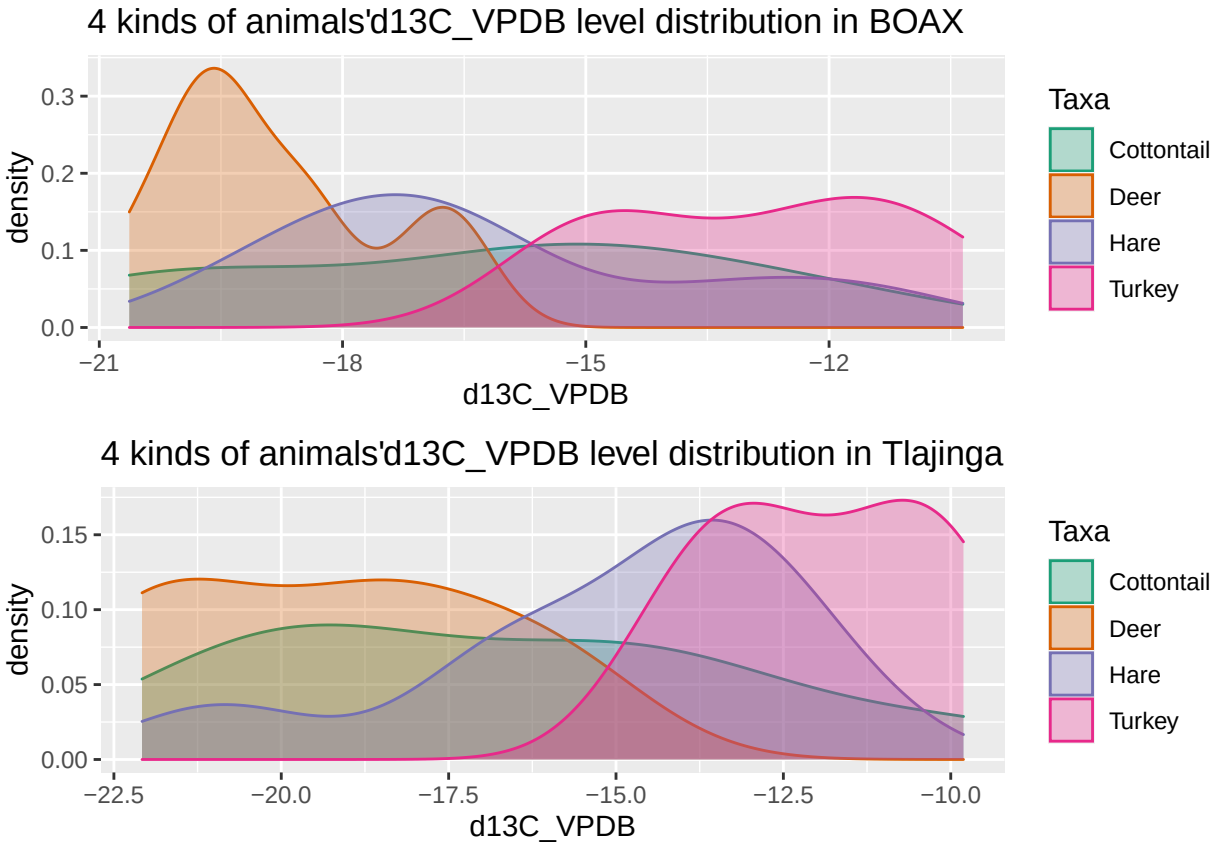
outlier falls on mean +- 2sd, we believe it is not a outlier.

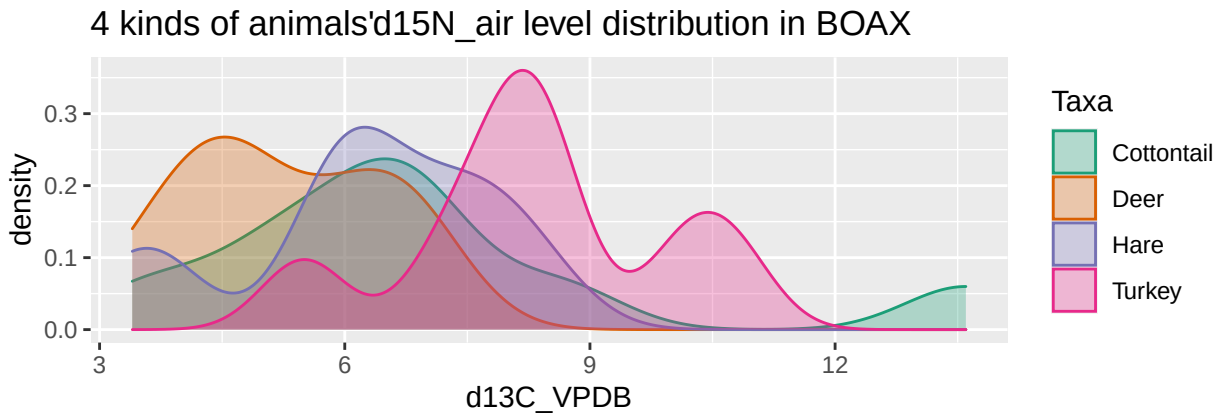
3.2 Testing

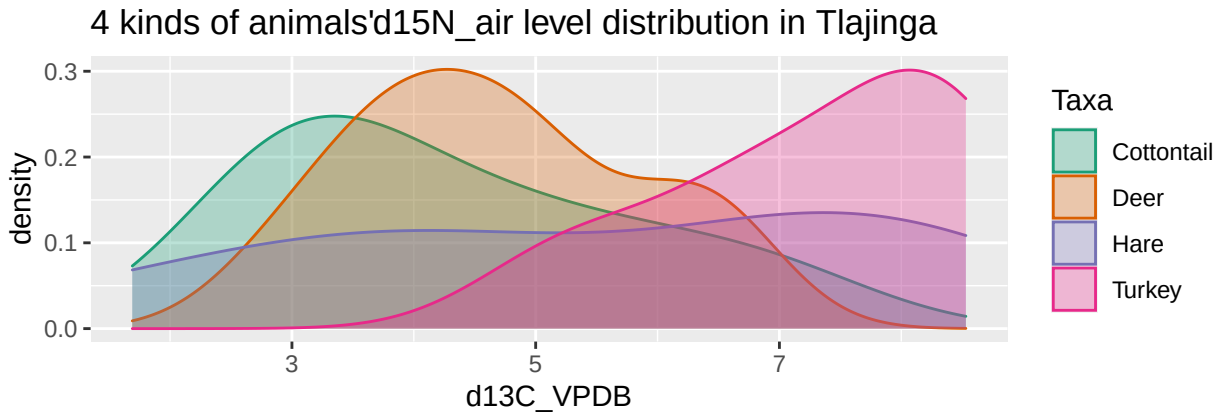
Density Test

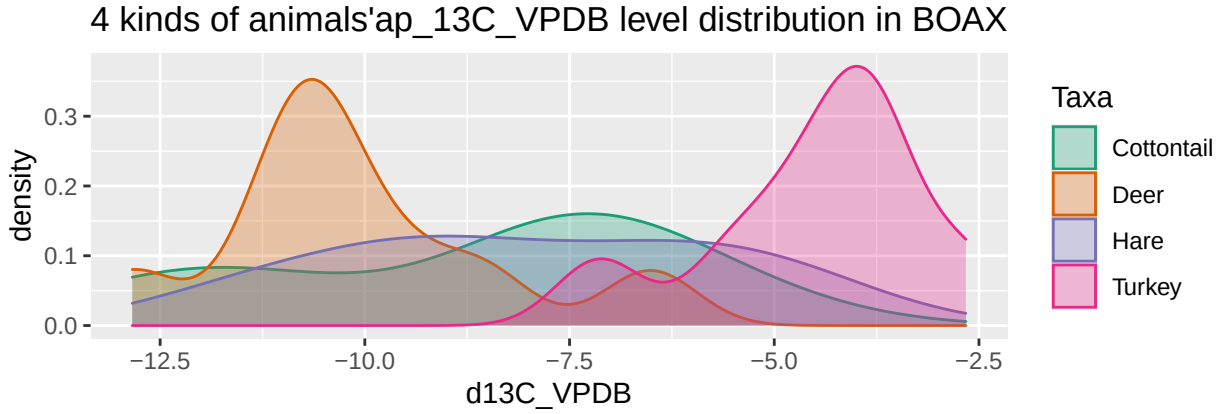
We would like to use the T-tests to find the difference of diets between two sites. We plan to try some statistic test or some regression to compare the isotopes level in two sites more precisely.

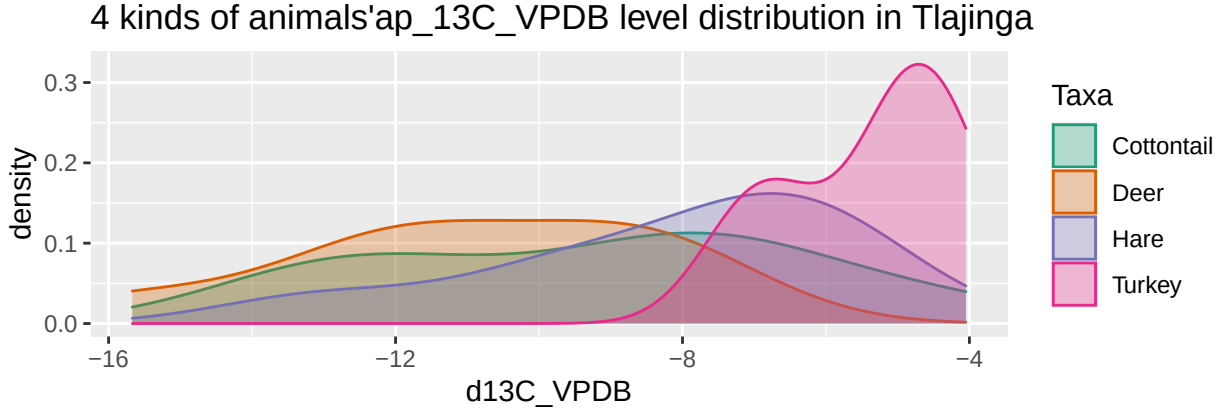
Firstly, we checked the normality.



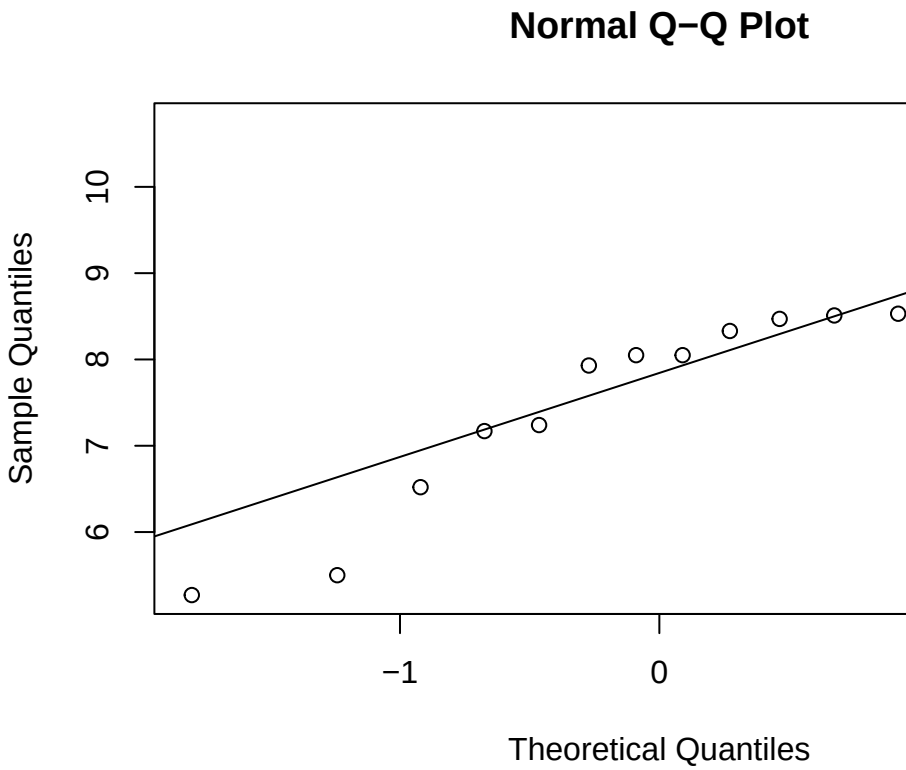

4 kinds of animals'd15N_air level distribution in BOAX
density
0.3
0.2
0.1
0.0
3
6
9
12
d13C_VPDB
Taxa
Cottontail
Deer
Hare
Turkey


4 kinds of animals'd15N_air level distribution in Tlajinga
density
0.3
0.2
0.1
0.0
3
5
7
d13C_VPDB
Taxa
Cottontail
Deer
Hare
Turkey


4 kinds of animals'ap_13C_VPDB level distribution in BOAX
density
0.3
0.2
0.1
0.0
−12.5
−10.0
−7.5
−5.0
−2.5
d13C_VPDB
Taxa
Cottontail
Deer
Hare
Turkey


4 kinds of animals'ap_13C_VPDB level distribution in Tlajinga
density
0.3
0.2
0.1
0.0
−16
−12
−8
−4
d13C_VPDB
Taxa
Cottontail
Deer
Hare
Turkey

- Normal Distribution Test (Q-Q Plot)



Take "Turkey" and "d15N_air" as an example:

Two-Sample T-tests

After we checked the distribution of isotope levels for each animals in each space, since the sample size is too small to identify if they follow normal distributions. We decided to do some shapiro.test.

Shapiro.test is a test of normality. If the p-value from the result is greater than the significance level (e.g. 0.05), we can assume the normality.

Two-Sample T-tests are as followings:

"d13C_VPDB" level

```
# Assumption
# H0 : v1  = v2
# Ha : v1 != v2
```

```
# "d13C_VPDB" level in turkey groups
test.turkey.d13C <- filter(aggdata, Taxa == "Turkey" & variable== "d13C_VPDB")
with(test.turkey.d13C, shapiro.test(value[Site == "BAOX"])) # p = 0.4606
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "BAOX"]
## W = 0.9237, p-value = 0.4606
```

```
with(test.turkey.d13C, shapiro.test(value[Site == "Tlajinga"])) # p = 0.7027
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  value[Site == "Tlajinga"]
## W = 0.94538, p-value = 0.7027
```

```
# From the output, the two p-values are greater than the significance level 0.05 implying that the dist

t.test(value ~ Site, data = test.turkey.d13C, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  value by Site
## t = -1.0102, df = 12, p-value = 0.3323
## alternative hypothesis: true difference in means between group BAOX and group Tlajinga is not equal
## 95 percent confidence interval:
##  -3.113316  1.140816
## sample estimates:
##     mean in group BAOX mean in group Tlajinga
##              -12.89625              -11.91000
```

```
# p = 0.3323
# Not reject!
```

```
# "d13C_VPDB" level in Cottontail groups
test.Cottontail.d13C <- filter(aggdata, Taxa == "Cottontail" & variable== "d13C_VPDB")
with(test.Cottontail.d13C, shapiro.test(value[Site == "BAOX"])) # p = 0.5969
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "BAOX"]
## W = 0.94142, p-value = 0.5969
```

```
with(test.Cottontail.d13C, shapiro.test(value[Site == "Tlajinga"])) # p = 0.3941
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "Tlajinga"]
## W = 0.91546, p-value = 0.3941
```

```
# From the output, the two p-values are greater than the significance level 0.05 implying that the dist
t.test(value ~ Site, data = test.Cottontail.d13C, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  value by Site
## t = 0.23657, df = 15, p-value = 0.8162
## alternative hypothesis: true difference in means between group BAOX and group Tlajinga is not equal
## 95 percent confidence interval:
##  -3.187167  3.983000
## sample estimates:
##     mean in group BAOX mean in group Tlajinga
##              -16.26333              -16.66125
```

```
# p = 0.8162
# Not reject!
```

```r
# "d13C_VPDB" level in Deer groups
test.Deer.d13C <- filter(aggdata, Taxa == "Deer" & variable== "d13C_VPDB")
with(test.Deer.d13C, shapiro.test(value[Site == "BAOX"])) # p = 0.3914
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "BAOX"]
## W = 0.9199, p-value = 0.3914
```

```r
with(test.Deer.d13C, shapiro.test(value[Site == "Tlajinga"])) # p = 0.2177
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "Tlajinga"]
## W = 0.89371, p-value = 0.2177
```

```r
# From the output, the two p-values are greater than the significance level 0.05 implying that the dist
t.test(value ~ Site, data = test.Deer.d13C, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  value by Site
## t = 0.44023, df = 16, p-value = 0.6657
## alternative hypothesis: true difference in means between group BAOX and group Tlajinga is not equal
## 95 percent confidence interval:
##  -1.581292  2.410181
## sample estimates:
##     mean in group BAOX mean in group Tlajinga
##              -18.83333              -19.24778
```

```r
# p = 0.6657
# Not reject!
```

```r
# "d13C_VPDB" level in Hare groups
test.Hare.d13C <- filter(aggdata, Taxa == "Hare" & variable== "d13C_VPDB")
with(test.Hare.d13C, shapiro.test(value[Site == "BAOX"])) # p = 0.2707
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "BAOX"]
## W = 0.91386, p-value = 0.2707
```

```r
with(test.Hare.d13C, shapiro.test(value[Site == "Tlajinga"])) # p = 0.2447
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "Tlajinga"]
## W = 0.89208, p-value = 0.2447
```

```r
# From the output, the two p-values are greater than the significance level 0.05 implying that the dist

t.test(value ~ Site, data = test.Hare.d13C, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  value by Site
## t = -0.86338, df = 17, p-value = 0.3999
## alternative hypothesis: true difference in means between group BAOX and group Tlajinga is not equal t
## 95 percent confidence interval:
##  -3.790379  1.589015
## sample estimates:
##     mean in group BAOX mean in group Tlajinga
##              -16.15818              -15.05750
```

```
# p = 0.3999
# Not reject!
```

"d15N_air" level

```
# "d15N_air" level in turkey groups
test.turkey.d15N <- filter(aggdata, Taxa == "Turkey" & variable== "d15N_air")
with(test.turkey.d15N, shapiro.test(value[Site == "BAOX"])) # p = 0.7717
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "BAOX"]
## W = 0.95604, p-value = 0.7717
```

```
with(test.turkey.d15N, shapiro.test(value[Site == "Tlajinga"])) # p = 0.4532
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "Tlajinga"]
## W = 0.91253, p-value = 0.4532
```

```
# From the output, the two p-values are greater than the significance level 0.05 implying that the dist

t.test(value ~ Site, data = test.turkey.d15N, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  value by Site
## t = 1.261, df = 12, p-value = 0.2313
## alternative hypothesis: true difference in means between group BAOX and group Tlajinga is not equal t
## 95 percent confidence interval:
##  -0.7365794  2.7607460
## sample estimates:
##     mean in group BAOX mean in group Tlajinga
##               8.323750               7.311667
```

```
# p = 0.2313
# Not reject!
```

```
# "d15N_air" level in Cottontail groups
test.Cottontail.d15N <- filter(aggdata, Taxa == "Cottontail" & variable== "d15N_air")
with(test.Cottontail.d15N, shapiro.test(value[Site == "BAOX"])) # p = 0.09437
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "BAOX"]
## W = 0.85937, p-value = 0.09437
```

```
with(test.Cottontail.d15N, shapiro.test(value[Site == "Tlajinga"])) # p = 0.2703
```

```
##
##  Shapiro-Wilk normality test
##
## data:  value[Site == "Tlajinga"]
## W = 0.89679, p-value = 0.2703
```

```
# From the output, the two p-values are greater than the significance level 0.05 implying that the dist
t.test(value ~ Site, data = test.Cottontail.d15N, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  value by Site
## t = 2.3623, df = 15, p-value = 0.0321
## alternative hypothesis: true difference in means between group BAOX and group Tlajinga is not equal
## 95 percent confidence interval:
##  0.2625433 5.1096789
## sample estimates:
##     mean in group BAOX mean in group Tlajinga
##               6.991111               4.305000
```

```
# p = 0.0321 < 0.05
# Reject!
```

Mann Whitney U Test

Mann-Whitney U Test for deer.

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 43, p-value = 0.8633
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 41, p-value = 0.673
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  value by Site
## W = 51.5, p-value = 0.3536
## alternative hypothesis: true location shift is not equal to 0
```

Mann-Whitney U Test for Cottontail.

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 38, p-value = 0.8884
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 43, p-value = 0.5414
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 60, p-value = 0.02065
## alternative hypothesis: true location shift is not equal to 0
```

Mann-Whitney U Test for Hare.

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 32, p-value = 0.3511
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  value by Site
## W = 45, p-value = 0.9671
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 50, p-value = 0.6574
## alternative hypothesis: true location shift is not equal to 0
```

Mann-Whitney U Test for Turkey.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  value by Site
## W = 15.5, p-value = 0.3012
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  value by Site
## W = 36, p-value = 0.1419
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  value by Site
## W = 32.5, p-value = 0.3012
## alternative hypothesis: true location shift is not equal to 0
```

```
#waining mean:
#  The impact of ties means the Wilcoxon rank sum distribution cannot be used to calculate exact p-valu


#Since this p-value is not less than 0.05, we fail to reject the null hypothesis.


# Mann-Whitney U Test for Cottontail.

# W=52, p-value= 0.03792

# we can reject the null hypothesis:
# two sites animals are the same
```

Check Mann-Whitney U Test with theory

```
Utable<-read.xlsx("Mann-Whitney_Table_0.01.xlsx")

Utest<-function(set1,set2){
  l1<-length(set1)
  l2<-length(set2)
  set<-c(set1,set2)
  set<-rbind(set,c(rep(1,l1),rep(2,l2)))
  rset<-rank(set[1,])
  R1<-sum(rset[1:l1])
  R2<-sum(rset[(l1+1):(l1+l2)])
  U1<-l1*l2+l1*(l1+1)/2-R1
  U2<-l1*l2+l2*(l2+1)/2-R2
  print(min(U1,U2))
  print(Utable[l1,l2])
  min(U1,U2)<Utable[l1,l2]
}
```

Mann-Whitney U Test for deer.

```
## [1] 38
## [1] 11
```

```
## [1] FALSE
```

```
## [1] 31
## [1] 9
```

```
## [1] FALSE
```

```
## [1] 29.5
## [1] 11
```

```
## [1] FALSE
```

Mann-Whitney U Test for Cottontail.

```
## [1] 34
## [1] 9
```

```
## [1] FALSE
```

```
## [1] 29
## [1] 9
```

```
## [1] FALSE
```

```
## [1] 12
## [1] 9
```

```
## [1] FALSE
```

Mann-Whitney U Test for Hare.

```
## [1] 32
## [1] 13
```

```
## [1] FALSE
```

```
## [1] 43
## [1] 13
```

```
## [1] FALSE
```

```
## [1] 38
## [1] 13
```

```
## [1] FALSE
```

Mann-Whitney U Test for Turkey.

```
## [1] 15.5
## [1] 4
```

```
## [1] FALSE
```

```
## [1] 12
## [1] 4
```

```
## [1] FALSE
```

```
## [1] 15.5
## [1] 4
```

```
## [1] FALSE
```

Because we find that Cottontail of d15N_air test reject the null hypothesis due to p-value. We want to check Mann-Whitney U Test without R package again.

Alpha=0.01, test statistic(7) and our critical value(7) cannot reject

Alpha=0.05. Since our test statistic (13) is greater than our critical value (7), we fail to reject the null hypothesis.

Theory : https://www.statology.org/mann-whitney-u-test/ Table : https://www.statology.org/mann-whitney-u-table/

4 Discussion

There exists an outlier point in the data set.

We can not make sure that they are same at 100% level. We can only say that at 95% confidence interval, the diets between BAOX and PATT are the same.

5 Conclusion

Citation

Appendix

Resource

multiple testing

Megan Goldman, 2008, Statistics for Bioinformatics