

Dina Sar Sam

Complete the following tables by hand

- For the following tables, fill in all the missing entries by hand.
- Show your work. This may mean adding pages.
- Post the final version back on canvas in one pdf document.

BETA DIVERSITY

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$

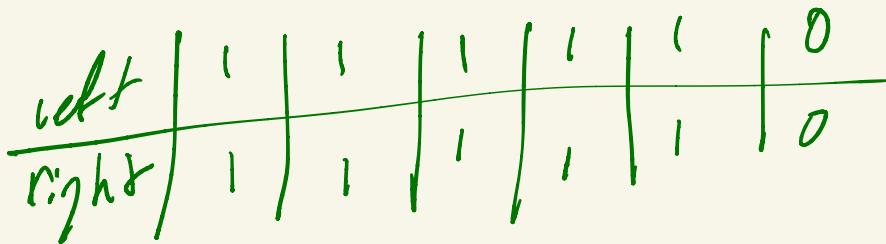
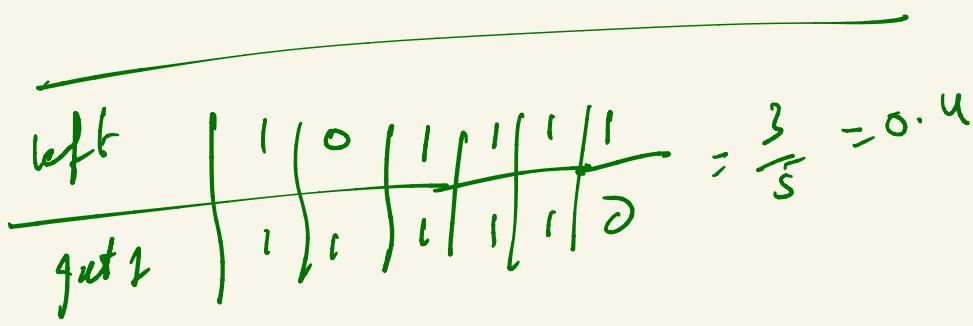
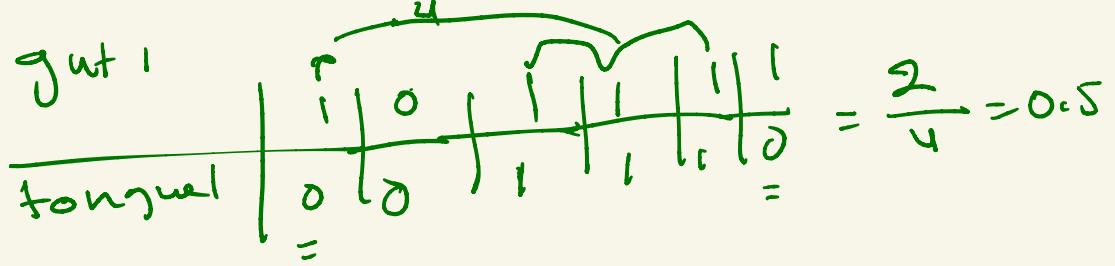
FeatureTable[Frequency]

	feature1	feature2	feature3	feature4	feature5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tonguel	0	0	87	12	0

DistanceMatrix

	gut1	left.palm1	right.palm1	tonguel
gut1	0.0	0.4	0.4	0.5
left.palm1	0.4	0.0	0.0	0.5
right.palm1	0.4	0.0	0.0	0.5
tonguel	0.5	0.5	0.5	0.0

symmetrical table



$$0/n = 1$$

By following the same methods
 I found the rest.

\rightarrow sum of different / sum

BETA DIVERSITY

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]

	feature1	feature2	feature3	feature4	feature5
gut1	42	0	37	99	1
left.palm1	12	1	22	88	0
right.palm1	25	3	23	86	0
tonguel	0	0	87	12	0
.

$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

X_{iA} : frequency of feature i in sample A

DistanceMatrix

	gut1	left.palm1	right.palm1	tonguel
gut1	0.0	0.19	0.15	0.65
left.palm1	0.19	0.0	0.07	0.69
right.palm1	0.15	0.07	0.0	0.70
tonguel	0.65	0.69	0.70	0.0

\rightarrow Solution steps

$$BC(A, B) = \frac{\sum_i |X_i:A - X_i:B|}{\sum_i (X_i:A + X_i:B)}$$

gut 1 and left, p, q, m 2, same as left patient and gut 1

Feature 1 Feature 2 Feature 3 Feature 4 Features

$$\frac{\sum |42-12|}{\sum |42+12|} + \frac{\sum |0-11|}{\sum |0+11|} + \frac{\sum |37-22|}{\sum |37+22|} + \frac{\sum |99-88|}{\sum |99+88|} + \frac{\sum |1-0|}{\sum |1+0|}$$

$$= \frac{30 + 1 + 15 + 11 + 1}{54 + 1 + 59 + 187 + 1} = \frac{58}{302} = 0.19$$

gut 1 and right, same as right patient and gut

$$\frac{\sum |42-25|}{\sum |42+25|} + \frac{\sum |0-3|}{\sum |0+3|} + \frac{\sum |37-23|}{\sum |37+23|} + \frac{\sum |99-86|}{\sum |99+86|} + \frac{\sum |1-0|}{\sum |1+0|}$$

$$= \frac{17 + 3 + 14 + 13 + 1}{67 + 3 + 60 + 185 + 1} = \frac{48}{366} = 0.13$$

gut 1 and tongue, same tongue and gut 1

~~$$\frac{\sum |42-0|}{\sum |42+0|} + \frac{\sum |0-0|}{\sum |0+0|} + \frac{\sum |37-87|}{\sum |37+87|} + \frac{\sum |99-12|}{\sum |99+12|} + \frac{\sum |1-0|}{\sum |1+0|}$$~~

$$= \frac{42/1}{42} + \frac{50}{124} + \frac{87}{111} + \frac{1}{1} = \frac{139}{237} \approx 0.59$$

left palm and tongue, same as tongue and left palm

$$\frac{\sum |12-0|}{\sum |12+0|} + \frac{\sum |1-0|}{\sum |1+0|} + \frac{\sum |22-87|}{\sum |22+87|} + \frac{\sum |88-12|}{\sum |88+12|} + \frac{\sum |0-0|}{\sum |0+0|}$$

$$\frac{12}{12} + \frac{1}{1} + \frac{65}{109} + \frac{76}{100} + \frac{0}{0} = \frac{154}{922} = \boxed{0.69}$$

left palm and right palm same as right palm and left palm

$$\frac{\sum |12-25|}{\sum |12+25|} + \frac{\sum |11-3|}{\sum |1+3|} + \frac{\sum |22-23|}{\sum |22+23|} + \frac{\sum |88-86|}{\sum |88+86|} + \frac{\sum |0-0|}{\sum |0+0|}$$

$$\frac{73}{37} + \frac{2}{4} + \frac{1}{45} + \frac{2}{174} = \frac{78}{260} = \underline{\underline{0.07}}$$

right palm and tongue same as tongue and right palm

$$\frac{\sum |25-0|}{\sum |25+0|} + \frac{\sum |3-0|}{\sum |3+0|} + \frac{\sum |23-27|}{\sum |23+27|} + \frac{\sum |86-12|}{\sum |86+12|} + \frac{\sum |0-0|}{\sum |0+0|}$$

$$\frac{25}{25} + \frac{3}{3} + \frac{64}{110} + \frac{74}{98} = \frac{166}{236} = \boxed{0.70}$$

UniFrac, UU unweighted → doesn't care about tree values.

Unweighted UniFrac distance:

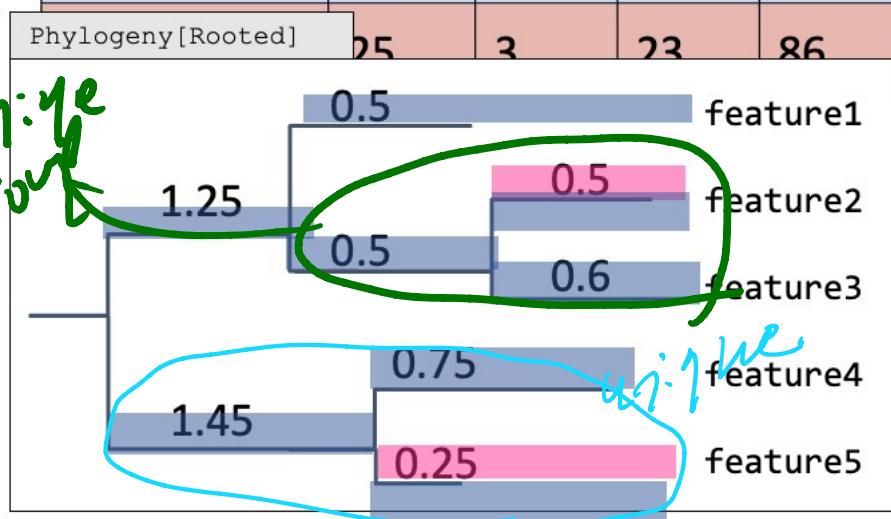
Applying this concept
I found the distance.

$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$

FeatureTable [Frequency]		feature1	feature2	feature3	feature4	feature5
		42	0	37	99	1
gut1		42	0	37	99	1
left.palm1		12	1	22	88	0

$$UU(\text{gut1}, \text{leftpalm1}) = (0.5 + 0.25) / (0.5 + 1.25 + 0.5 + 0.5 + 0.6 + 1.45 + 0.75 + 0.25)$$

$$UU(\text{gut1}, \text{leftpalm1}) = 0.75 / 5.8$$



		gut1	left.palm1	right.palm1	tongue1
0	0.0			0.14	
0		0.13	0.0	0.0	
		0.13	0.0	0.0	
		0.14	0.18	0.68	
				0.0	

Basically studying the unique groups

if 1 and more \rightarrow 60 set \rightarrow geometric spaces
 GM mean \rightarrow varying number within sample
 log-ratio

Complete **clr-transform** by hand for sample 1 & 2

Show work

ID	sam1	sam2	sam3	sam4	sam5	sam6	sam7
pseudomonas	40	1000	50	30	2	80	18
bacillus	10	647	68	150	2025	100	5
escherichia	5	208	12	70	200	98	30

feature
 GM mean

GM =
 wt%
 \downarrow
 $\ln \left(\frac{w_1}{\sum w_i} \right)$ \rightarrow extend \rightarrow and with them.

Calculation

Geometric mean for Both sample 1 and 2

for sample 1

$$40 \times 10 \times 5 = 2000 \rightarrow \text{Cub root } \sqrt[3]{2000} \\ = \boxed{12.6}$$

for sample 2

$$1000 \times 647 \times 208 = 134776000$$

$$\sqrt[3]{13477600} \Rightarrow 512.954$$

ID	Sample 1	Sample 2	ln Sample 1	ln Sample 2
Pseudomonas	$40/12.6$ $= 3.1$	$1000/512.954$ $= 1.9$	$\ln(3.1)$ $= 1.16$	$\ln(1.9)$ $= 0.7$
bacillus	$10/12.6$ $= 0.79$	$647/512.954$ $= 1.3$	$\ln(0.79)$ $= -0.230$	$\ln(1.3)$ $= 0.232$
Escherichia	$5/12.6$ $= 0.4$	$208/512.954$ $= 0.40$	$\ln(12.6)$ $= -0.924$	$\ln(0.4)$ $= -0.902$



taking natural
log for
each