

Contingency Table

	A	B	C	D	E	F	G	H	I	J	K	L
				P(x1)P(x2)	MI	Jaccard Index		Category	Proportion	InfectedObserved	InfectedEpected	
	0	0		0	#DIV/0!	100		G1	0.5	50	99.5	
	1	0		0	#DIV/0!			G2	0.5	21	99.5	
	0	0		0	#DIV/0!				71	149		
	0	0		0	#DIV/0!			pValue	1.02727E-08			
	1	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							
	0	0		0	#DIV/0!							
	1	0		0	#DIV/0!							

Formula = $I(x_1, x_2) = P(x_1, x_2) \log (P(x_1, x_2) / P(x_1)P(x_2))$

Jaccard Index

Number in both sets = 2

Number in either set = 2

$$(2/2) * 100 = 100$$

Jaccard Index = 100

Chi-Squared X2

$$X^2 = \sum_{i=1}^k ((O_i - e_i)^2) / e_i$$

The code in excel is CHISQ.TEST(I4:J4,K2:K3) or (71:149,99.5,99.5)

pValue/Chi-S = 1.02727E-08

P2

2.1A:

α Level = 0.5

	C	D	E	F
105		Correlation r	0.380875036	
166		N:	2400	
21		TS:	20.17161871	
146		DF:	2398	
162		P Value	1.04095E-83	
159				

The correlation that we were able to gather shows the value has very little to do with each other due to the negative association value. The P-value also seems to be lower than the alpha which would result in the rejection of the null hypothesis. The association between the data clearly tells me the genes have no significant to each other. From the data gathered.

2.2B:

α Level = 0.5

Coefficient r:	0.931219633
N:	110
TS:	26.55314238
DF:	108
Pvalue	3.73732E-49

The data that was given was able to give us a correlation that is positive in the terms of association. The P-value was also lower than the alpha level and that would result in the rejection of the null hypothesis. The association between the data clearly show the genes have a very high association with each other and the graph also shows it seeing how they move upward in slope.

2.3C:

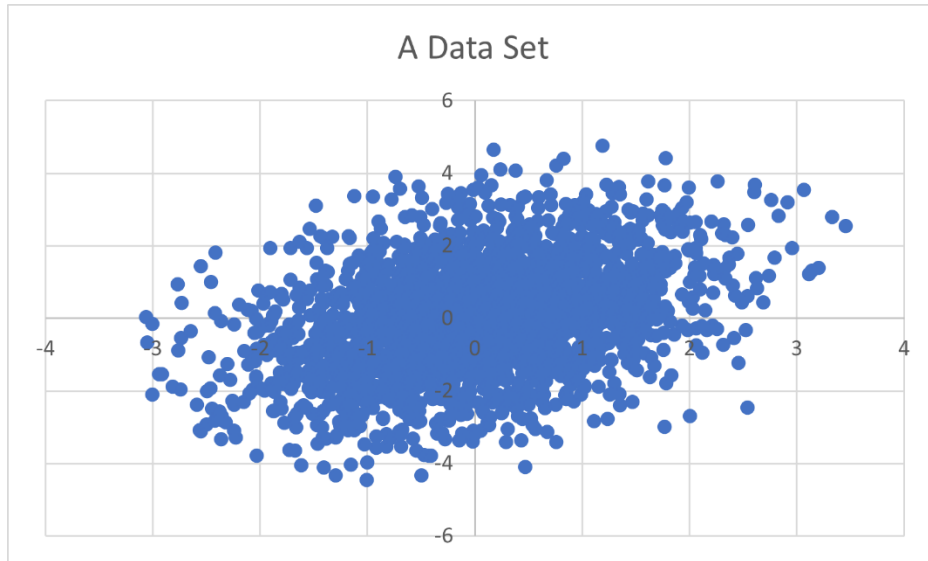
α Level = 0.5

Correlation r:	0.041179
N:	2100
TS:	1.887761
DF:	2098
Pvalue:	0.059196

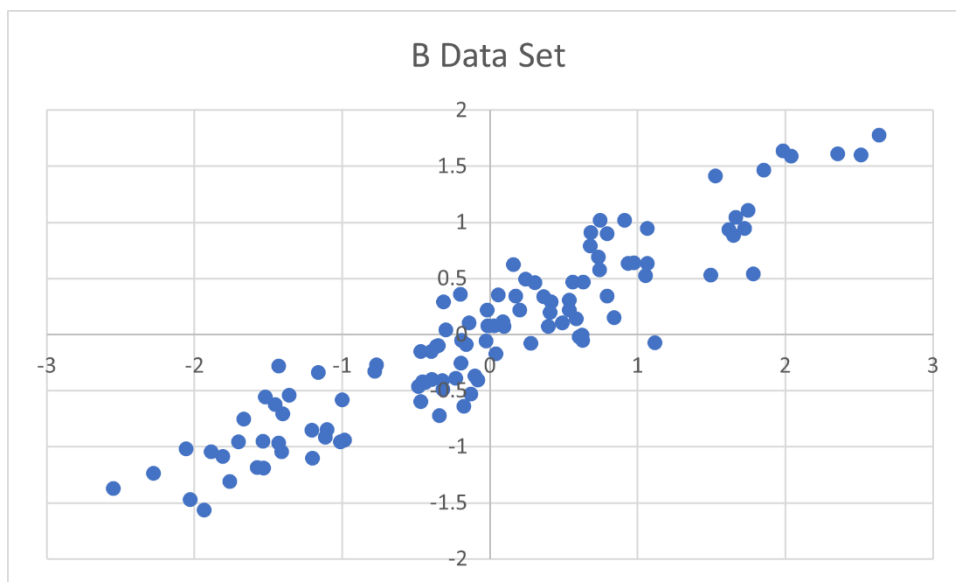
The data that was given was able to give us a correlation that is negative, but very little it's near the center in terms of association. The P-Value was also higher than the Alpha level which means we can't reject the null hypothesis. The association between the data clearly shows the genes have some association to each other the graph shows that.

2.4D:

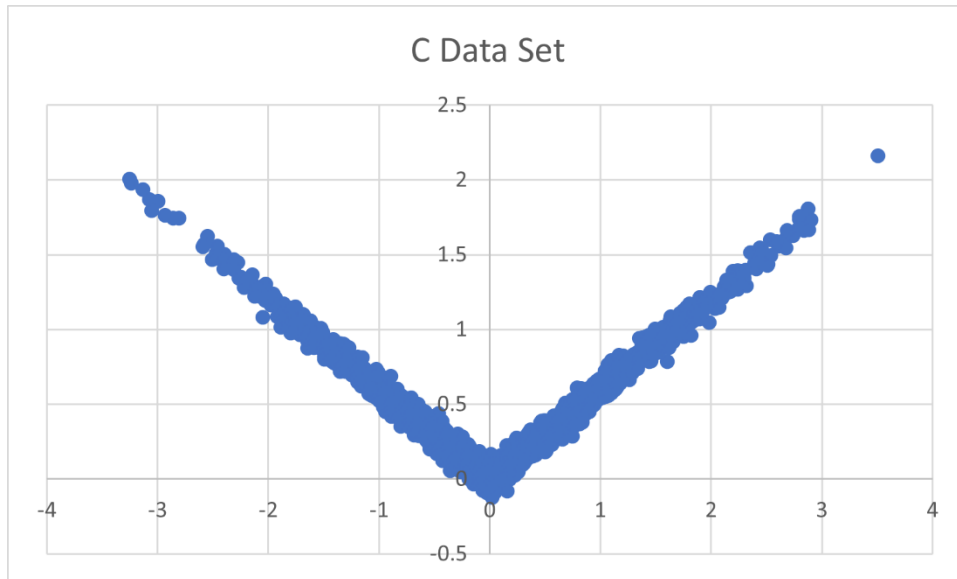
A



B



C



2.5E:

B data set gives the higher correlation with the number of .93 which shows that data has a very high correlation with one another. The C data set has a higher P-Value which tells us that we can't reject the null hypothesis and giving us the higher association shown by the graph. The data set do show a good association with the correlation and P-Value. C to me seems the data set to have a higher association due to the weird V the graph is displaying.