Q1 (a): Descriptive Analysis

a	
Mean	1.518365421
Standard Error	0.000207596
Median	1.51768
Mode	1.5159
Standard Deviation	0.003036864
Sample Variance	9.22254E-06
Kurtosis	4.931737386
Skewness	1.625430506
Range	0.02278
Minimum	1.51115
Maximum	1.53393
Sum	324.9302
Count	214
Largest(1)	1.53393
Smallest(1)	1.51115
Confidence Level(95.0%)	0.000409205

b	
Mean	13.40785047
Standard Error	0.05582187
Median	13.3
Mode	13.02
Standard Deviation	0.816603556
Sample Variance	0.666841367
Kurtosis	3.052232409
Skewness	0.454181454
Range	6.65
Minimum	10.73
Maximum	17.38
Sum	2869.28
Count	214
Largest(1)	17.38
Smallest(1)	10.73
Confidence Level(95.0%)	0.110034054

c	
Mean	2.68453271
Standard Error	0.09860097
Median	3.48
Mode	0
Standard Deviation	1.442407845
Sample Variance	2.080540391
Kurtosis	-0.410318963
Skewness	-1.152559318
Range	4.49
Minimum	0
Maximum	4.49
Sum	574.49
Count	214
Largest(1)	4.49
Smallest(1)	0
Confidence Level(95.0%)	0.194358672

d	
Mean	1.444906542
Standard Error	0.03412937
Median	1.36
Mode	1.54
Standard Deviation	0.499269646
Sample Variance	0.249270179
Kurtosis	2.060568969
Skewness	0.907289809
Range	3.21
Minimum	0.29
Maximum	3.5
Sum	309.21
Count	214
Largest(1)	3.5
Smallest(1)	0.29
Confidence Level(95.0%)	0.067274582

e	
Mean	72.65093458
Standard Error	0.052946861
Median	72.79
Mode	72.86
Standard Deviation	0.774545795
Sample Variance	0.599921188
Kurtosis	2.967902956
Skewness	-0.730447225
Range	5.6
Minimum	69.81
Maximum	75.41
Sum	15547.3
Count	214
Largest(1)	75.41
Smallest(1)	69.81
Confidence Level(95.0%)	0.104366939

f	
Mean	0.497056075
Standard Error	0.044582917
Median	0.555
Mode	0
Standard Deviation	0.652191846
Sample Variance	0.425354203
Kurtosis	54.68969853
Skewness	6.55164831
Range	6.21
Minimum	0
Maximum	6.21
Sum	106.37
Count	214
Largest(1)	6.21
Smallest(1)	0
Confidence Level(95.0%)	0.087880235

g	
Mean	8.956962617
Standard Error	0.09728477
Median	8.6
Mode	8.43
Standard Deviation	1.423153487
Sample Variance	2.025365848
Kurtosis	6.681977951
Skewness	2.047053913
Range	10.76
Minimum	5.43
Maximum	16.19
Sum	1916.79
Count	214
Largest(1)	16.19
Smallest(1)	5.43
Confidence Level(95.0%)	0.191764224

h	
Mean	0.175046729
Standard Error	0.033989209
Median	0
Mode	0
Standard Deviation	0.497219261
Sample Variance	0.247226993
Kurtosis	12.54108358
Skewness	3.416424569
Range	3.15
Minimum	0
Maximum	3.15
Sum	37.46
Count	214
Largest(1)	3.15
Smallest(1)	0
Confidence Level(95.0%)	0.066998301

i	
Mean	0.057009346
Standard Error	0.006660772
Median	0
Mode	0
Standard Deviation	0.097438701
Sample Variance	0.0094943
Kurtosis	2.662015617
Skewness	1.75432747
Range	0.51
Minimum	0
Maximum	0.51
Sum	12.2
Count	214
Largest(1)	0.51
Smallest(1)	0
Confidence Level(95.0%)	0.013129474

Q1(a): Anova and Correlation

Table 1: ANOVA Single Factor for all additives

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
a	214	324.9302	1.518365	9.22E-06
b	214	2869.28	13.40785	0.666841
c	214	574.49	2.684533	2.08054
d	214	309.21	1.444907	0.24927
e	214	15547.3	72.65093	0.599921
f	214	106.37	0.497056	0.425354
g	214	1916.79	8.956963	2.025366
h	214	37.46	0.175047	0.247227
i	214	12.2	0.057009	0.009494

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	943261.1	8	117907.6	168332	0	1.943226
Within Groups	1342.757	1917	0.700447			
Total	944603.8	1925				

Summary

H_o: $\mu_a = \mu_b = \mu_c = \mu_d = \mu_e = \mu_f = \mu_g = \mu_h = \mu_i$

H_1: At least 1 inequality among all groups

α used in ANOVA is 0.05, since *P*-value is < 0.05. We reject the null hypothesis in favor of the alternate hypothesis.

Table 2: R-Correlation for all additives

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
a	1								
b	-0.11929	1							
c	-0.17111	-0.23121	1						
d	-0.41637	0.16095	-0.50441	1					
e	-0.52997	-0.19453	0.07295	-0.03984	1				
f	-0.33062	-0.25369	0.03037	0.373052	-0.23838	1			
g	0.802771	-0.22748	-0.50943	-0.27401	-0.14608	-0.35034	1		
h	0.072309	0.275346	-0.46788	0.49675	-0.2702	-0.00967	-0.06504	1	
i	0.142833	-0.35926	0.03542	-0.02473	0.007477	0.030933	0.135707	-0.03326	1

Summary

It can be concluded that only some of the additives are related to others whilst some of the additives are not related at all to the other additives. Additive *a* and *g* are highly correlated with each other as shown in the correlation above as well proven in the cluster study in figure 1. Other additive as well correlate with each other (highlighted in yellow) but it is not as significant as *a* and *g*. This data concludes that, additive *a* is directly proportional to additive *g*. Both this additive has to be added at the same amount. To add to that, both additive *c* and *h* has significant influence as well on at least 3 other additives.

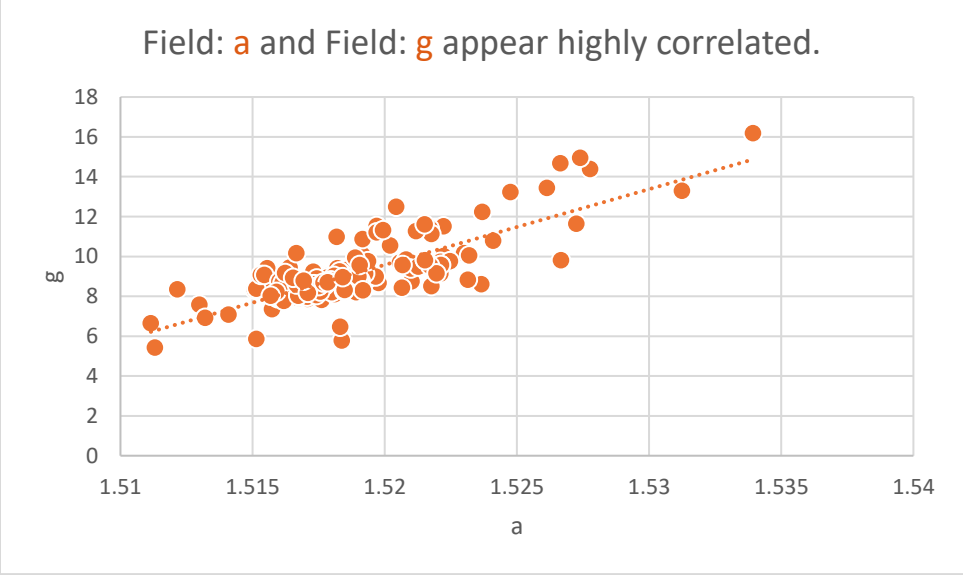


Figure 1: Cluster of g and a

Q1 (b): Graphical Representation and distribution study

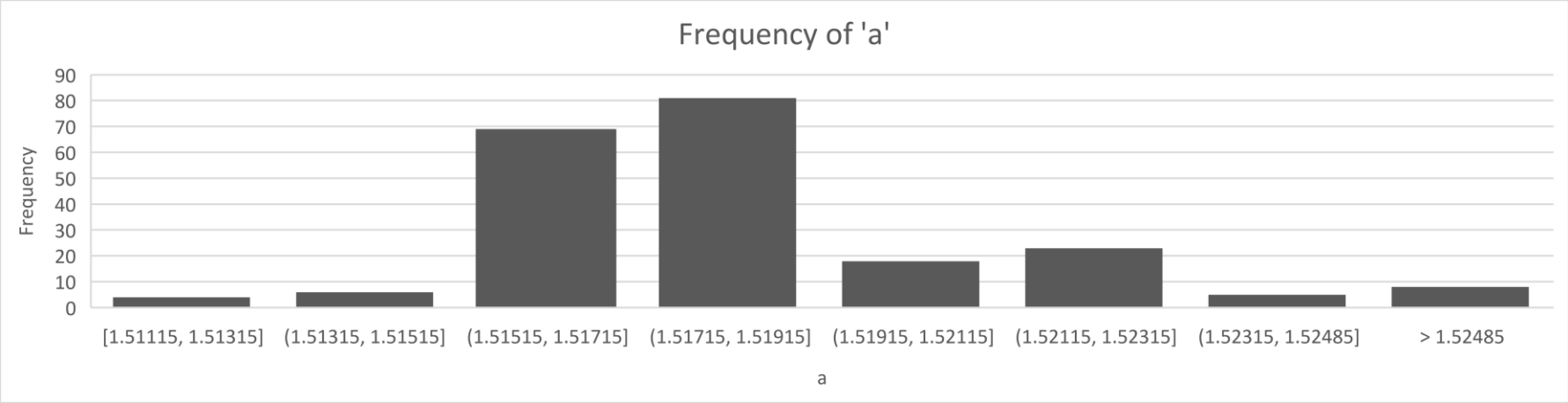


Figure 2: Bar Chart of frequency vs *a*

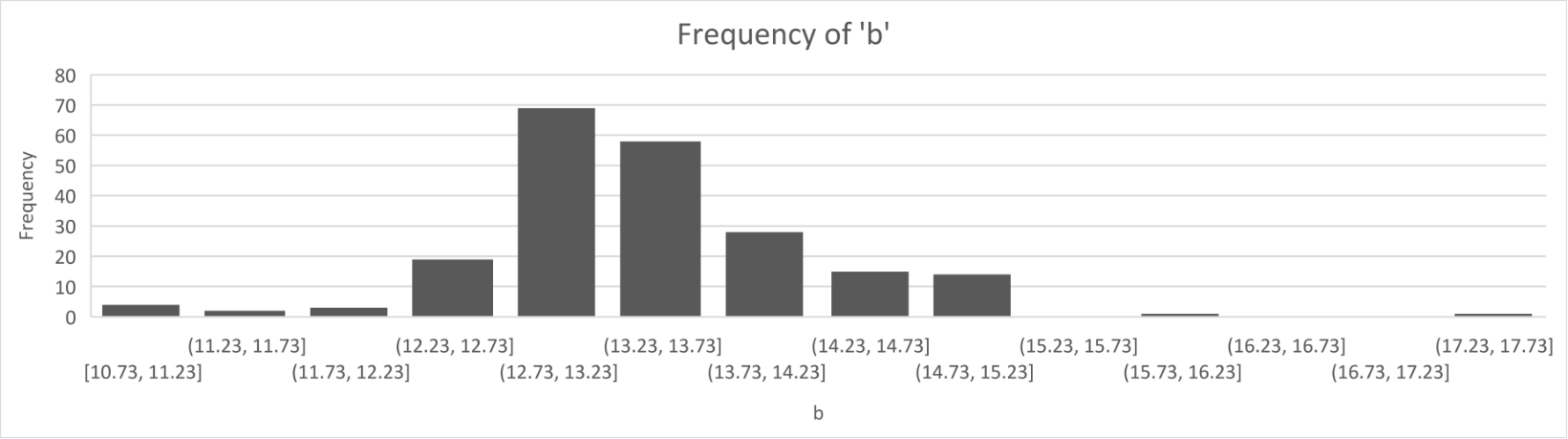


Figure 3: Bar Chart of frequency vs *b*

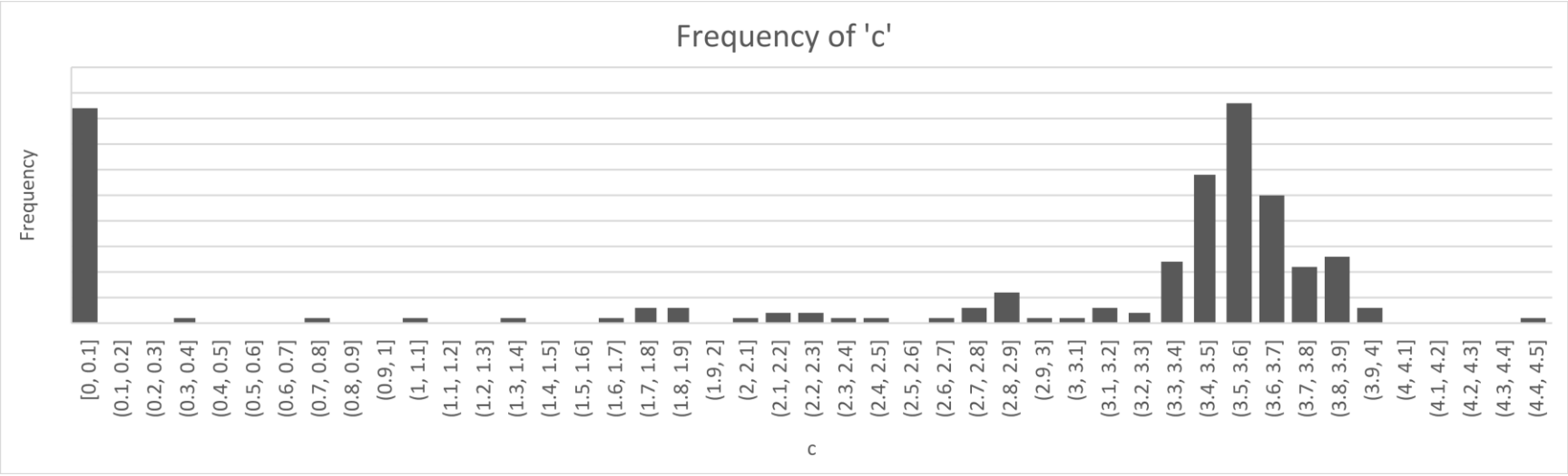


Figure 4:Bar Chart of frequency vs *c*

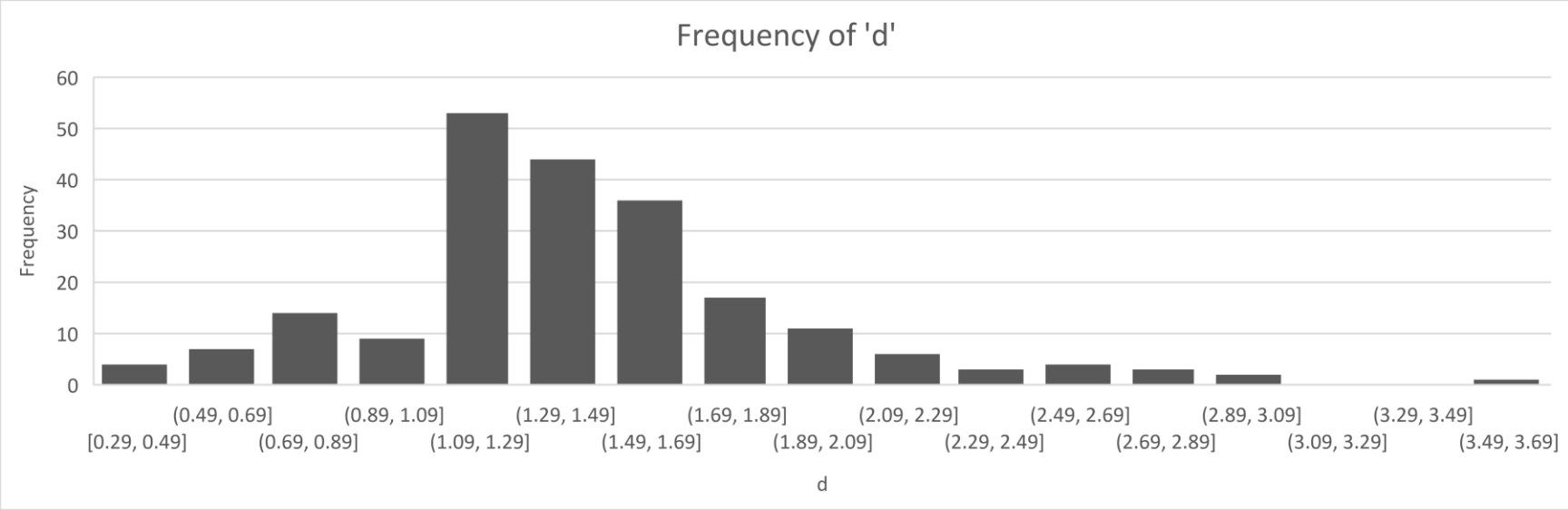


Figure 5:Bar Chart of frequency vs *d*

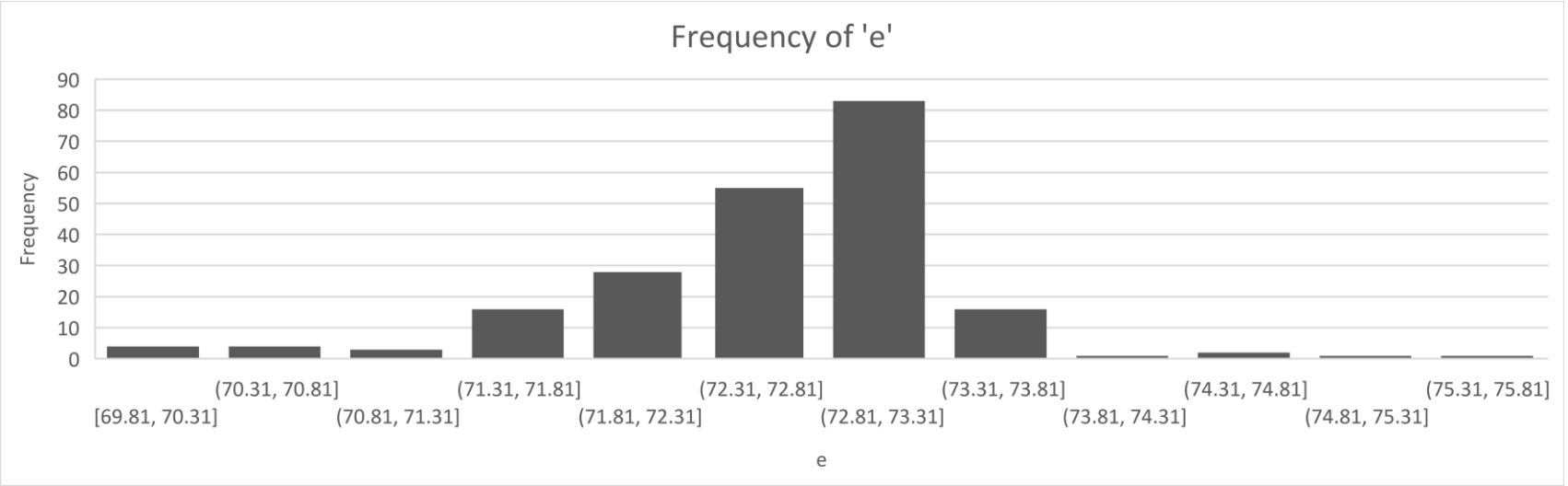


Figure 6: Bar Chart of frequency vs *e*

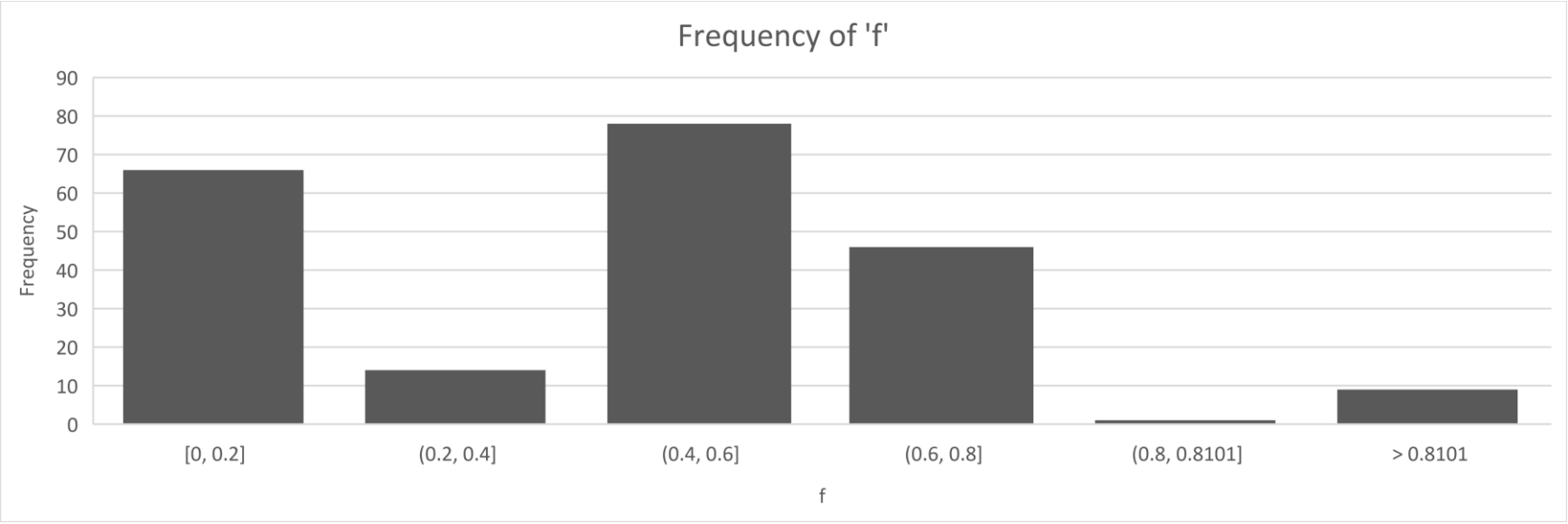


Figure 7: Bar Chart of frequency vs f

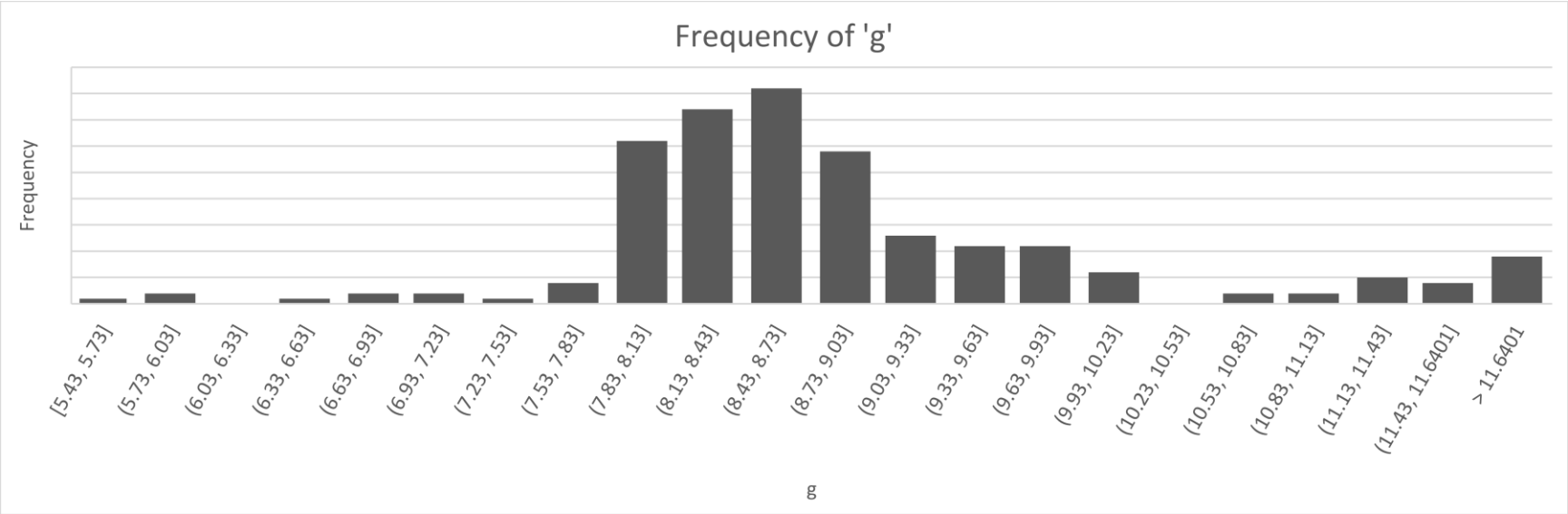


Figure 8: Bar Chart of frequency vs *g*

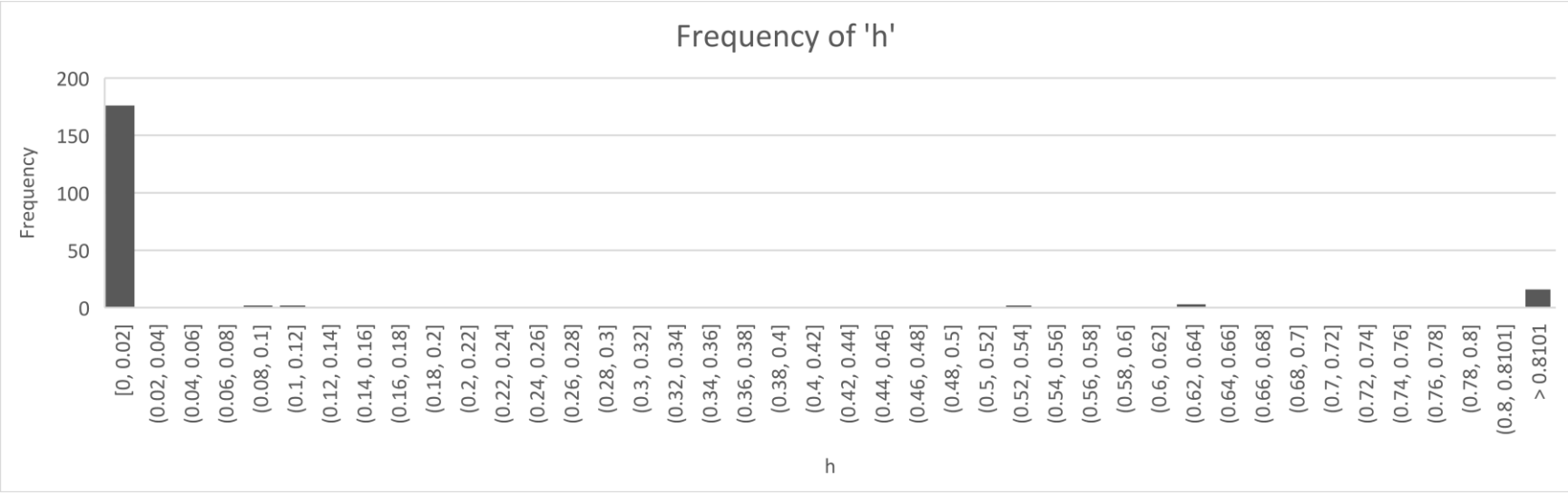


Figure 9: Bar Chart of frequency vs *h*

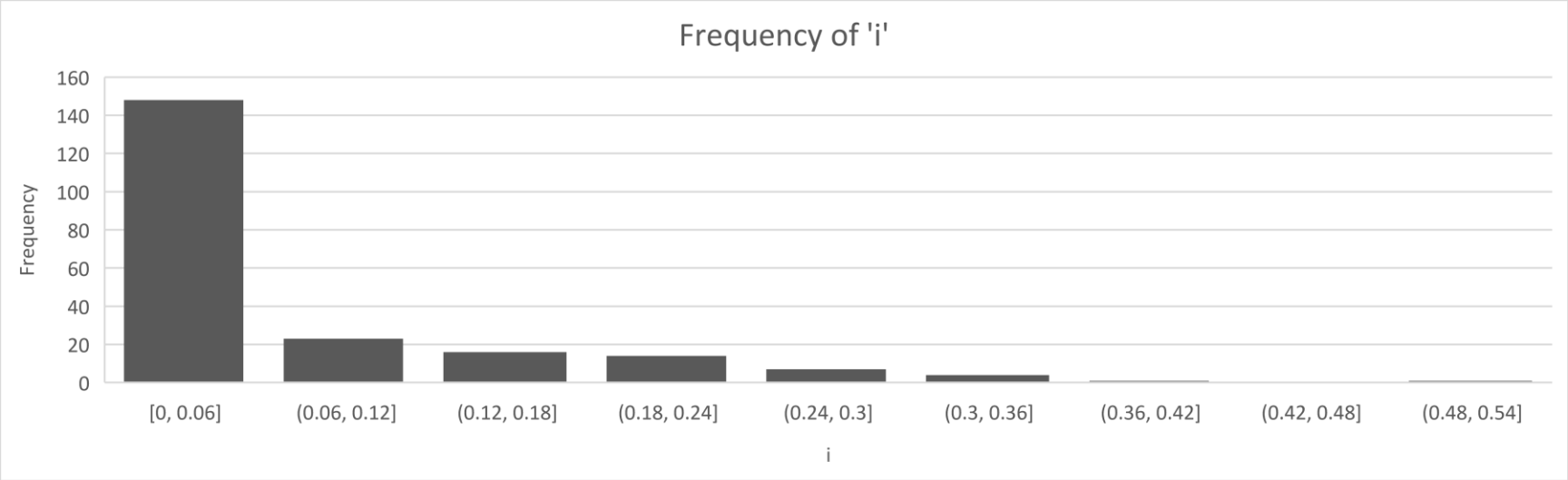


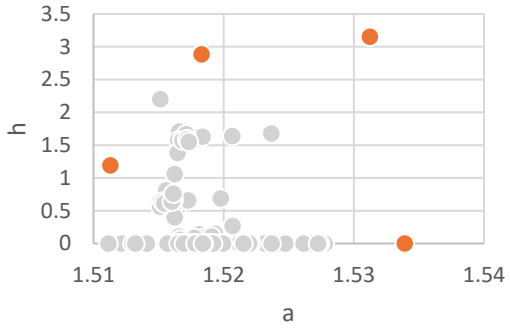
Figure 10: Bar Chart of frequency vs *i*

Summary

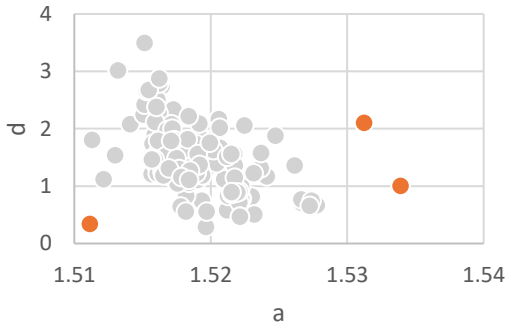
Based on figure 2 to 10, all additives have wide ranges of data. Additive *h* and *i* can be fairly concluded that it is added consistently. Whilst additive *f* is the most inconsistent among all the additives. Also, additive *c* can be concluded to be the added with the least accuracy, the frequency of outlier is high and seemingly same frequency as the one that is added within range. Other additives have its inconsistency but the data is still fairly reliable as the frequency of outliers are not as significant as the one within range.

Q1 (c): Clustering

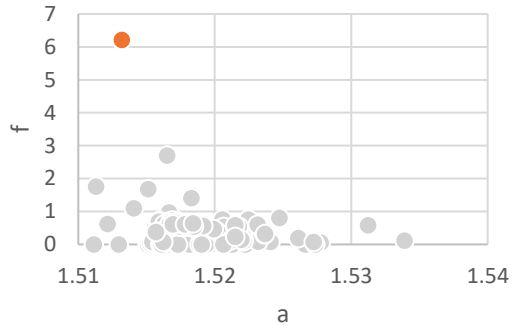
'a' and 'h' appear to form a cluster with 4 outliers.



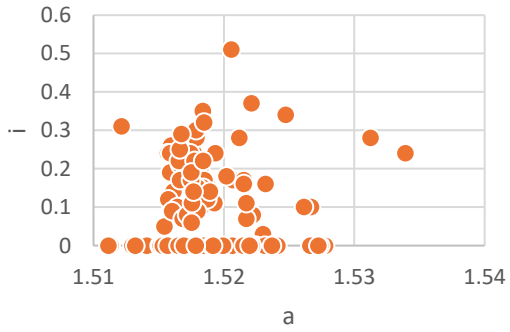
'a' and 'd' appear to form a cluster with 3 outliers.



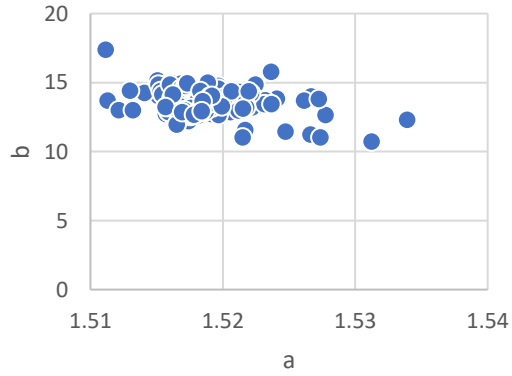
'a' and 'f' appear to form a cluster with 2 outliers.



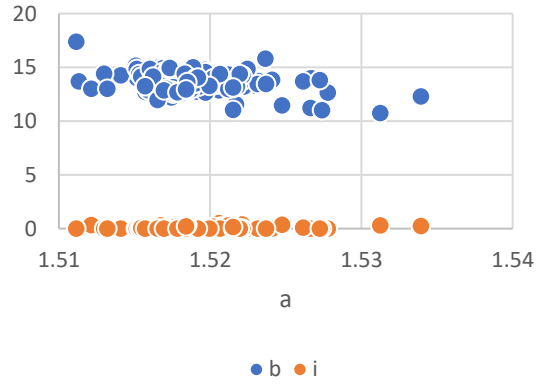
Field: i appears highly determined by Field: a.

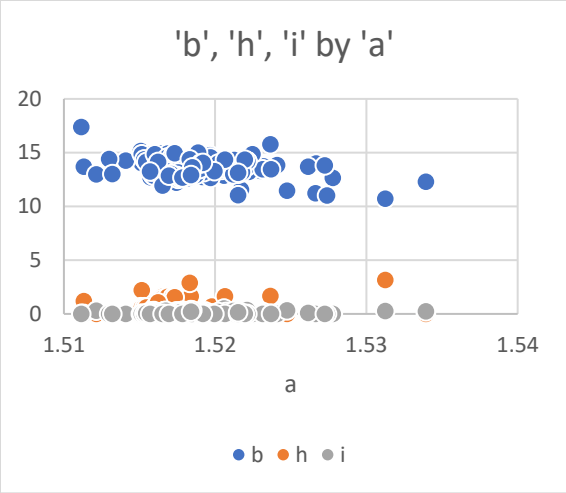


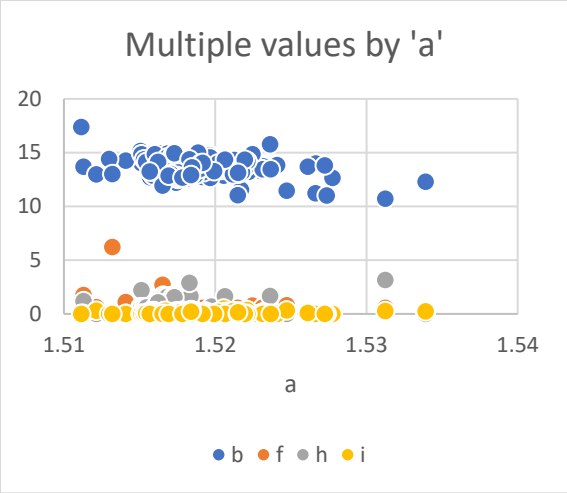
'b' by 'a'

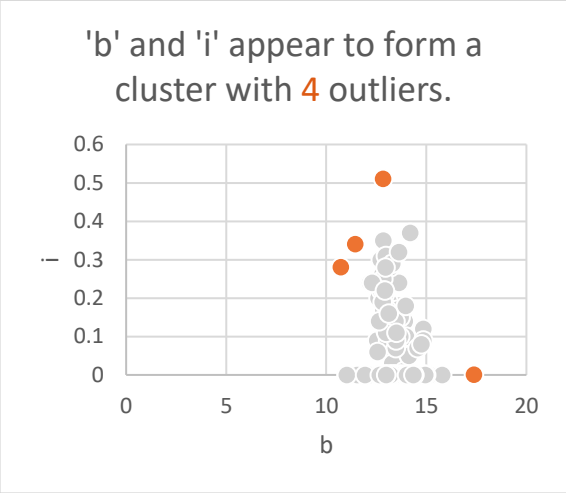


'b', 'i' by 'a'









Based on figures above, we can have a glance of the cluster with relation to *a*. Additives *h, d* and *f* produces 4, 3 and 2 outliers respectively. When *a* is relate to *i* and *b*, there is no outlier recorded. Thus, we can see from figure of multiple values by 'a', where multiple additives are clustered together with no outliers formation. This shows that outliers that are present during individual comparison with *a* are removed as they are part of the clusters of other additives. For example, in figure of cluster *a* and *h*, additives *h* has 4 outliers which is included in the additives *b* and *i* as shown in figure of cluster *b* and *i*.

Q2: R-Correlation

	SoilMoisture	*Average_Temp*	*Min_Temp*	*Max_Temp*	*Precipitation*	*Working_days*	*HA_Harvested*	*FFB_Yield*
SoilMoisture	1							
Average_Temp	-0.64987782	1						
Min_Temp	0.015838684	0.180396489	1					
Max_Temp	-0.499936248	0.761082633	-0.1247544	1				
Precipitation	-0.552000687	-0.369386365	0.345943849	-0.461117465	1			
Working_days	-0.057014648	0.076320657	0.068413506	-0.03911225	-0.12789702	1		
HA_Harvested	-0.326539455	0.446515045	0.024395877	-0.314827259	-0.265865829	0.048876344	1	
FFB_Yield	0.003182901	-0.005494353	0.103829694	-0.071200905	-0.289603724	0.116364072	-0.350221838	1

Based on R-Correlation, it can be deduced that FBB Yield is affected most by precipitation on the soil. It turns out that Precipitation is affected most by soil moisture, which in this case means that rain will keep the soil moist and in return increase the precipitation on the soil as we have more chemicals content from the rain this indirectly increases the FBB yield.