Report BE-303 Applied Biostatistics

Student name: - **Krishna Kumar Dixit**

Roll No: - B21013

---

1. Task 1

1.1. Selection of statistical test

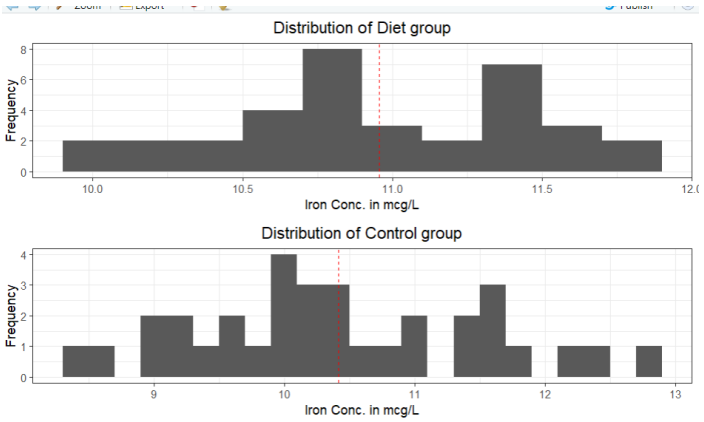
- Parametric tests

As both groups have ratio scaled variables -> Parametric tests
Samples were normally distributed , confirmed by qqplots , histograms and Shapiro-Wilk normality test

Results of Shapiro-Wilk normality test

data: df$Diet
W = 0.97246, p-value = 0.5145

data: df$Control
W = 0.97546, p-value = 0.6435



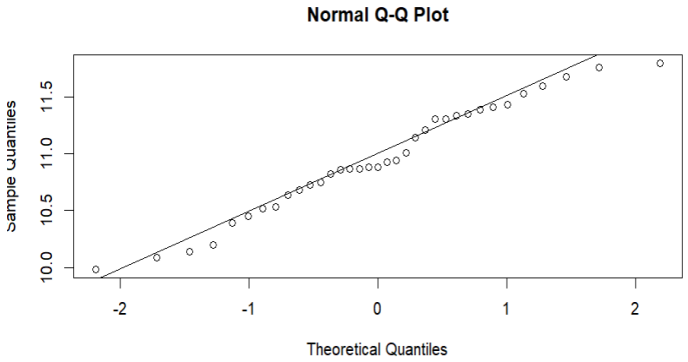
p-value for Shapiro-wilk test for both groups is higher than alpha threshold of 0.05 , Therefore , we failed to reject that distribution is not normal.

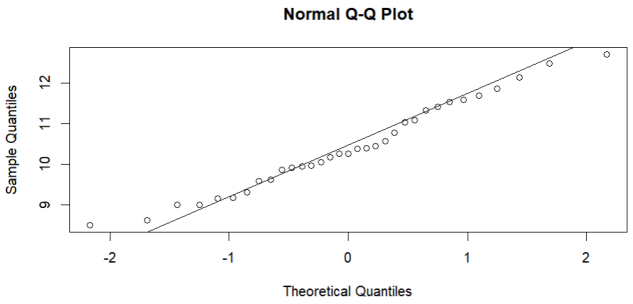
- Test for Variance similarity:-
  F-score test results in F-score of 5.17 which is greater than its standard value.Therefore , we reject the null hypothesis that there is no difference between the variances.

As there is a significant difference between the variances of two groups, the Welche test was chosen with best power .

1.2. Statistical analysis (including graphs)

cohen_d	0.633903160039002
control_std_dev	1.11122628822232
diet_std_dev	0.488493310379696
diff_mean	0.538493506493507
diff_std	0.622732977842628
F	5.17473092676789
mean_control	10.4163636363636
mean_diet	10.9548571428571
n_control	33
n_diet	35
ratio_of_var	5.17473092676789





Results from your statistical analysis including high quality graph

Figure 1: *The picture shows the as qqplot of Diet group on the left and Control on the right .*

1.3. Conclusions

- p_value =`.01402178` as the pvalue is less than alpha threshold of 0.05 <- null hypothesis is rejected
- With an alpha threshold of 5% It can be said that there is a statistically significant difference between Diet and Control groups. That means there is an increase in iron levels in diet groups in comparison to Control groups.

---

2. Task 2

2.1. Selection of statistical test

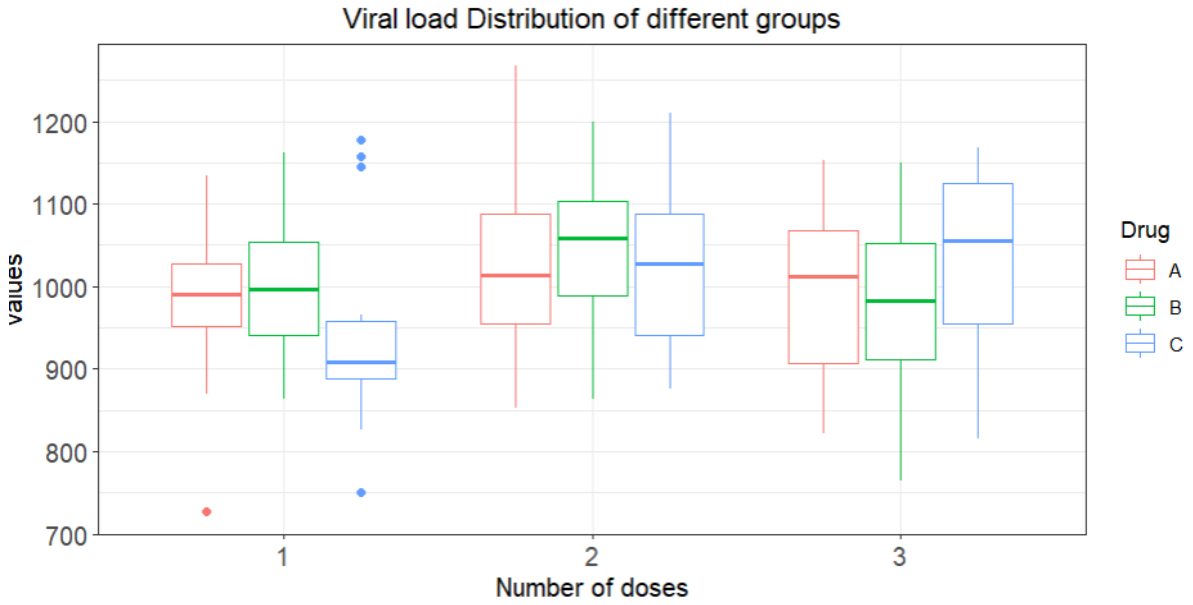
- Two-way anova is chosen for the analysis as independent or explanatory variables are categorical and the dependent variable is continuous.

Correlation matrix for the variables is shown on the right.
It shows there is no correlation between the different independent variables.

	S01	S02	S03	S04	S05	S06	S07	S08	S09
S01	1.0000000000	-0.102804205	-0.03383720	0.0005208752	0.020056317	0.40639102	0.02560249	0.450852807	0.1883093
S02	-0.1028042054	1.000000000	-0.13024079	0.3540499372	0.316896068	0.03147427	-0.23310663	-0.008810344	0.3977847
S03	-0.0338372006	-0.130240794	1.00000000	0.1200147298	-0.637402855	-0.09616557	0.23672782	0.164474857	-0.3065932
S04	0.0005208752	0.354049937	0.12001473	1.0000000000	-0.072736491	-0.41098078	0.13957354	0.368573686	0.1936770
S05	0.0200563169	0.316896068	-0.63740286	-0.0727364907	1.000000000	0.03001869	-0.13409130	0.004800226	0.1171647
S06	0.4063910194	0.031474275	-0.09616557	-0.4109807786	0.030018687	1.00000000	-0.30039237	0.014343452	0.1832633
S07	0.0256024854	-0.233106632	0.23672782	0.1395735410	-0.134091298	-0.30039237	1.00000000	0.308176140	-0.4811283
S08	0.4508528070	-0.008810344	0.16447486	0.3685736859	0.004800226	0.01434345	0.30817614	1.000000000	0.2003014
S09	0.1883093124	0.397784707	-0.30659316	0.1936769798	0.117164680	0.18326328	-0.48112834	0.200301415	1.0000000

Data is normally distributed.

2.2. Statistical analysis



Summary statistics:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Drug	2	3049	1524	0.136	0.8728	
Doses	2	58080	29040	2.594	0.0787	.
Drug:Doses	4	50193	12548	1.121	0.3497	
Residuals	126	1410530	11195			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

2.3. Conclusions

- F_Drug= 0.136< 3.07 Type of Drugs is not having any significant difference on viral load
- F_Doses = 2.594 < 3.07 Number of Doses are not having any significant difference on viral load
- F_DOses:Drugs = 1.121 < 2.44 => No relation between Drugs and doses
- Overall : There is no significant relation between any variables i.e Drug A = Drug B = Drug C and the Number of doses is also not having any significant difference in the inhibition of virus.

---

3. Task 3

3.1. Selection of statistical test

In this case we want to analyze if there is an effect in the number of T cells with Vitamin C levels

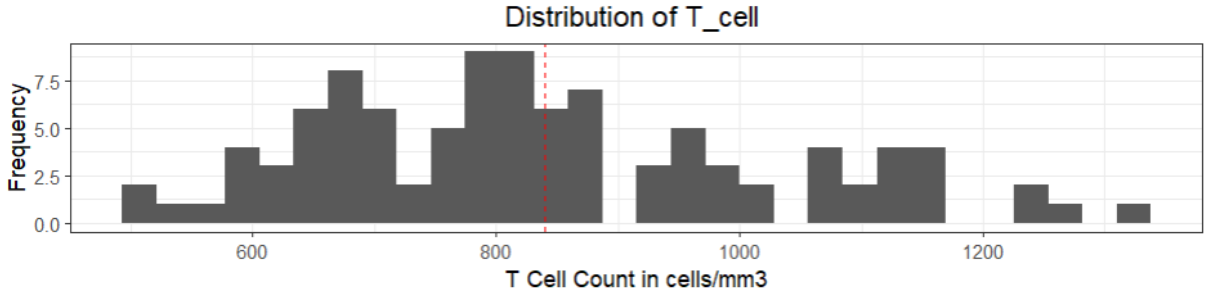
so, we will do correlation analysis. Both T_Cells and C_level are continuous variables

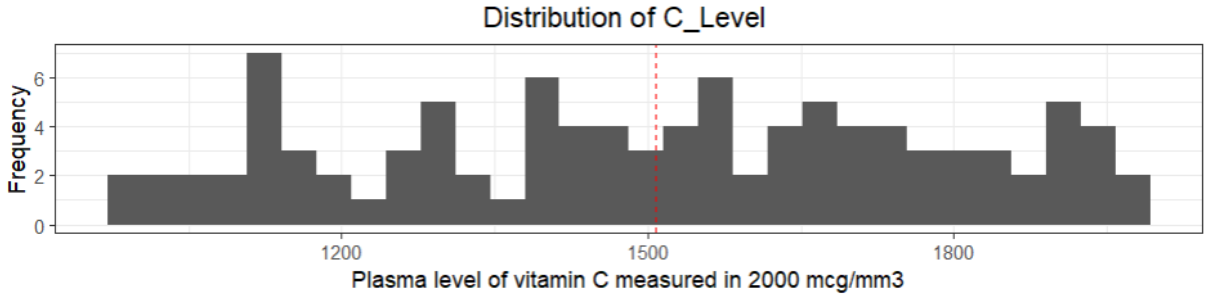
Correlation = -0.71401

As there is a high correlation between the variables, linear regression is the best fit model but there is some noise in the data.

3.2. Statistical analysis

Distribution of data:-

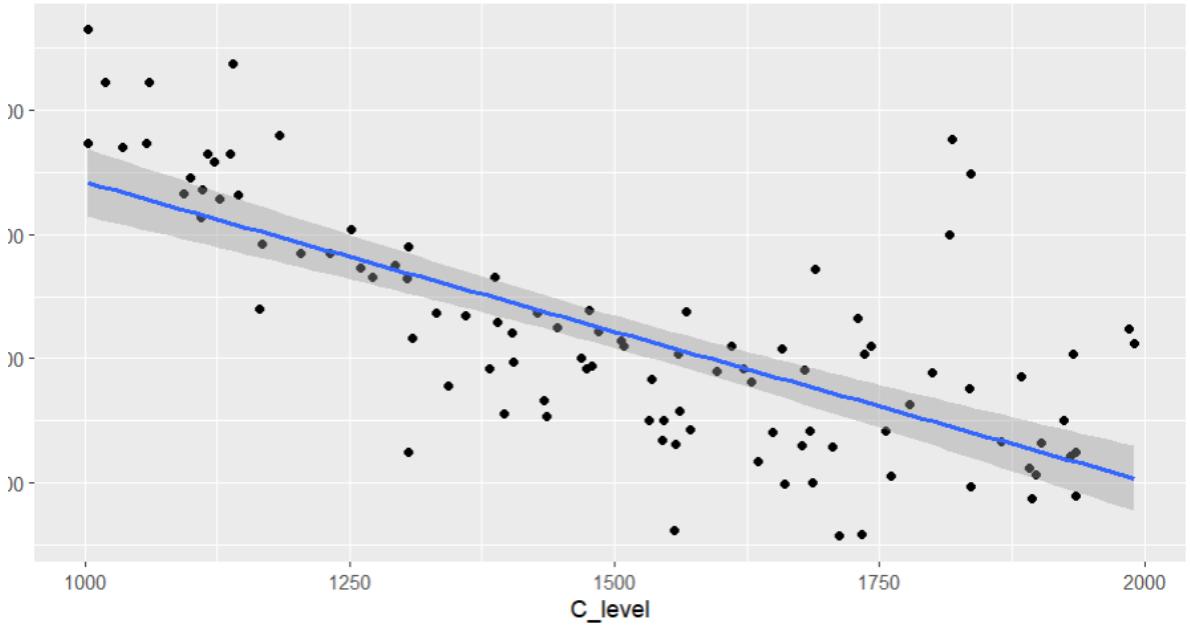




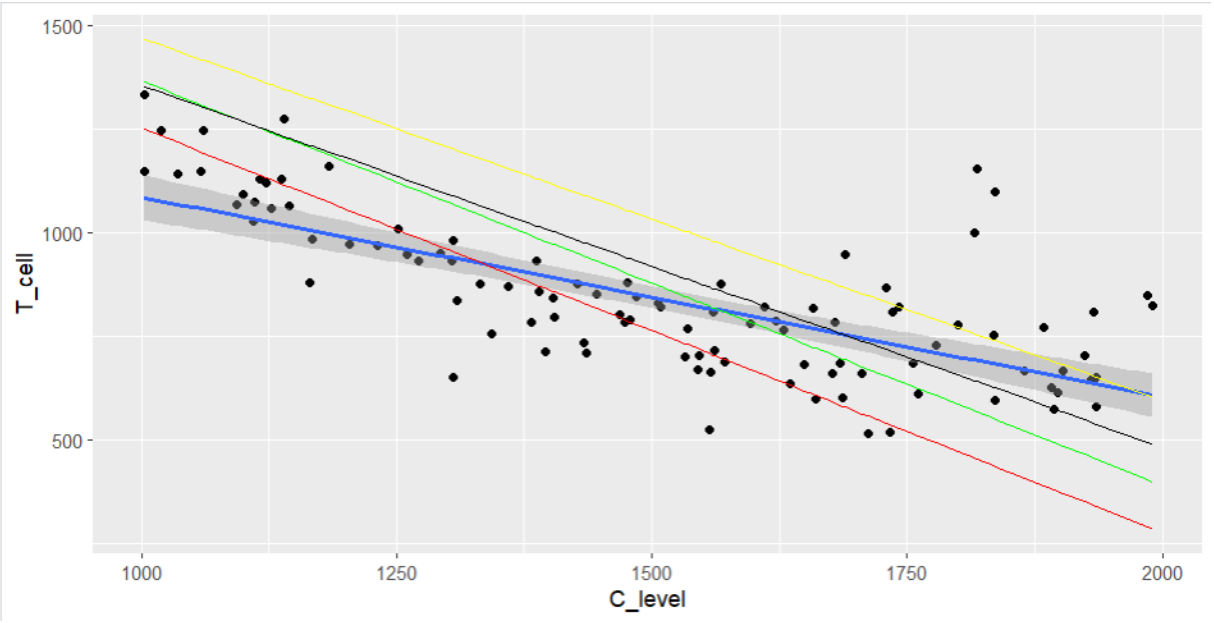
Statistical information

alpha	0.05
b0	2284.15536451789
b1	-0.924552565222639
CIb0_0	2227.15738631017
CIb0_1	2341.1533427256
CIb1_0	-0.9760088204031
CIb1_1	-0.873096310042178
corr	-0.714010177914988
mean_C_level	1507.448
mean_T_cell	840.09
ratio	2.19709459310472
reg_out	num [1:100] 1119 1257 867 889 444 ...
RSS	3434430.94091735
SEb0	28.7220523965878
SEb1	0.0259295031841204
t_quantile	1.98446745450848
var_C	76642.8306020202
var_T	34883.7190909091
x	num [1:100] 1261 1111 1533 1509 1990 ...
x_mean	840.09
x_xmean	num [1:100] 421 271 693 669 1150 ...
y	num [1:100] 946 1073 700 821 824 ...
y_mean	1507.448
y_ymean	num [1:100] -561 -434 -807 -686 -683 ...

Without CI -



With Confidence Intervals:-



3.3. Conclusions :-

- Data is biased and heteroscedastic.
- There is a negative correlation between Count of T_cell and Level of vitamin C . That means with higher values of Vitamin C levels the count of C_level is decreasing.