Assignment 5

Sainivedhitha Arunajatesan

Exercise 1

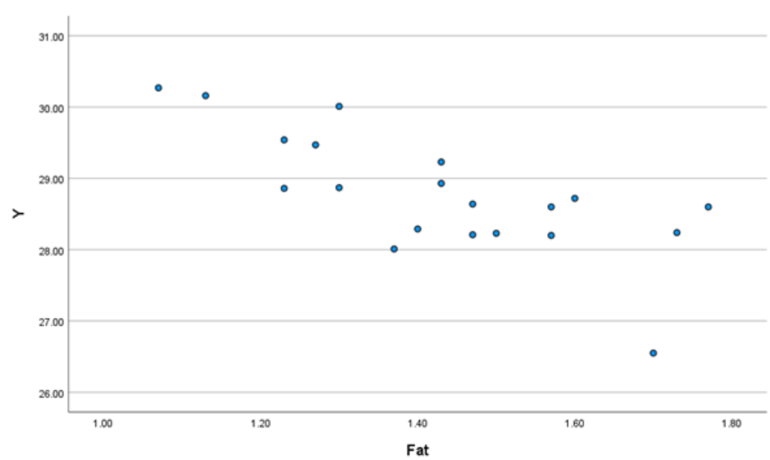
Regression:

My last name contains 12 letters. Hence, the data set R3 is used along with RATS data set.

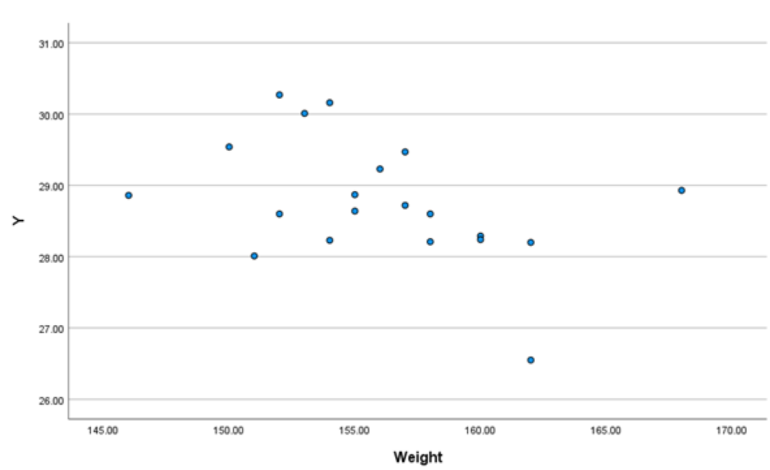
R3 data set

Scatter Plot:

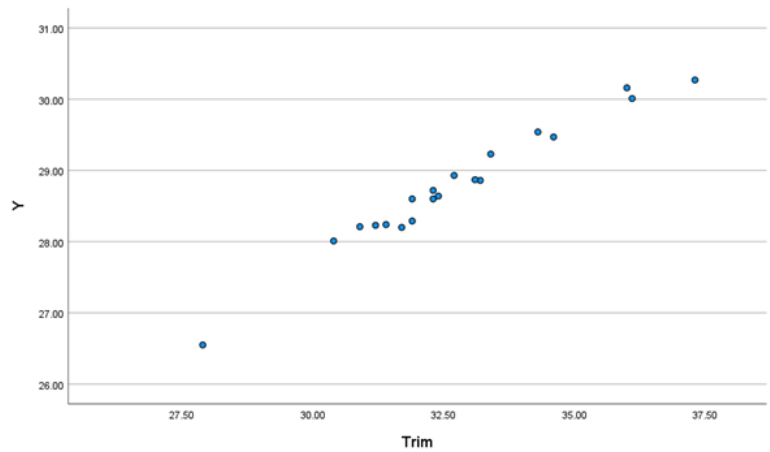
Y - Fat



Y - Weight



Y - Trim



Normality Test:

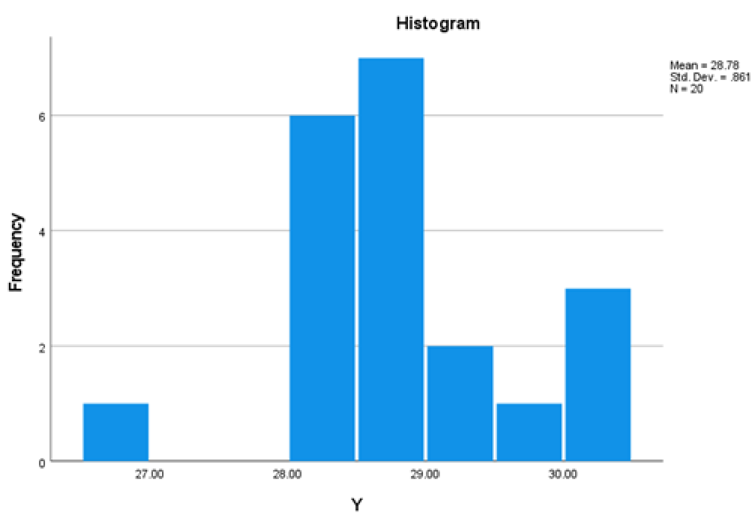
Y

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Y	.150	20	.200*	.936	20	.201

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



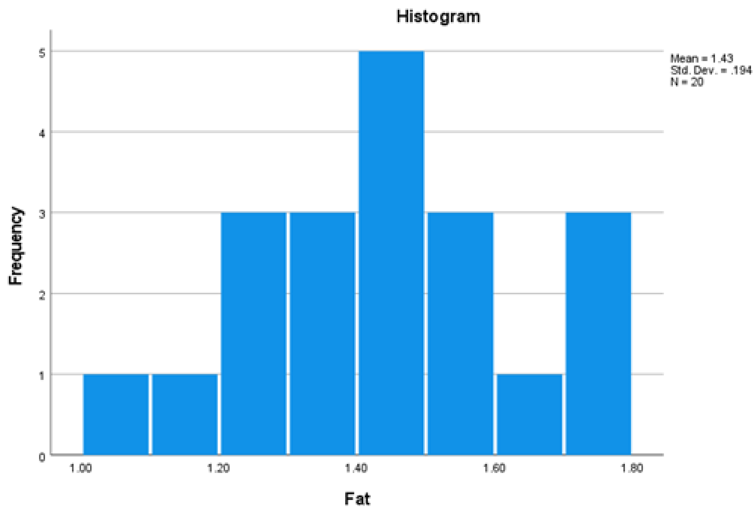
Fat

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Fat	.093	20	.200*	.980	20	.929

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



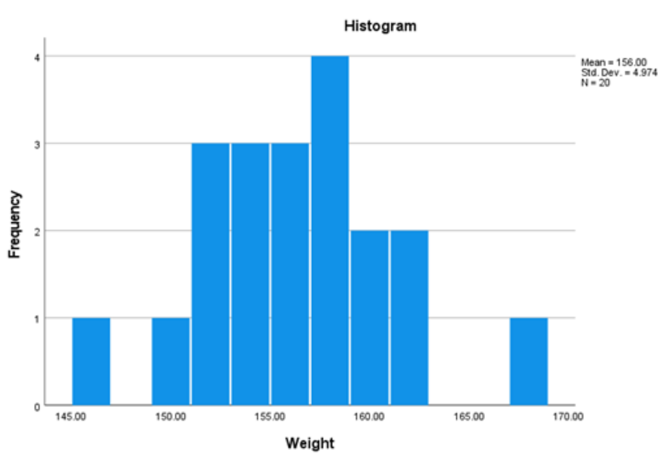
Weight

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight	.094	20	.200*	.983	20	.962

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



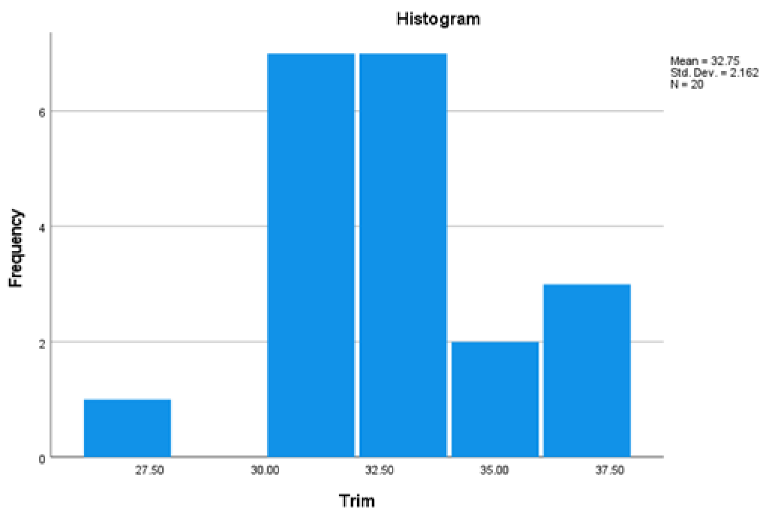
Trim

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Trim	.132	20	.200*	.963	20	.600

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Linear Regression:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	16.435	1.737		9.464	.000
	Fat	.110	.327	.025	.336	.741
	Weight	-.005	.009	-.028	-.525	.607
	Trim	.395	.027	.992	14.590	.000

a. Dependent Variable: Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.984[a]	.968	.962	.16721

a. Predictors: (Constant), Trim, Weight, Fat

b. Dependent Variable: Y

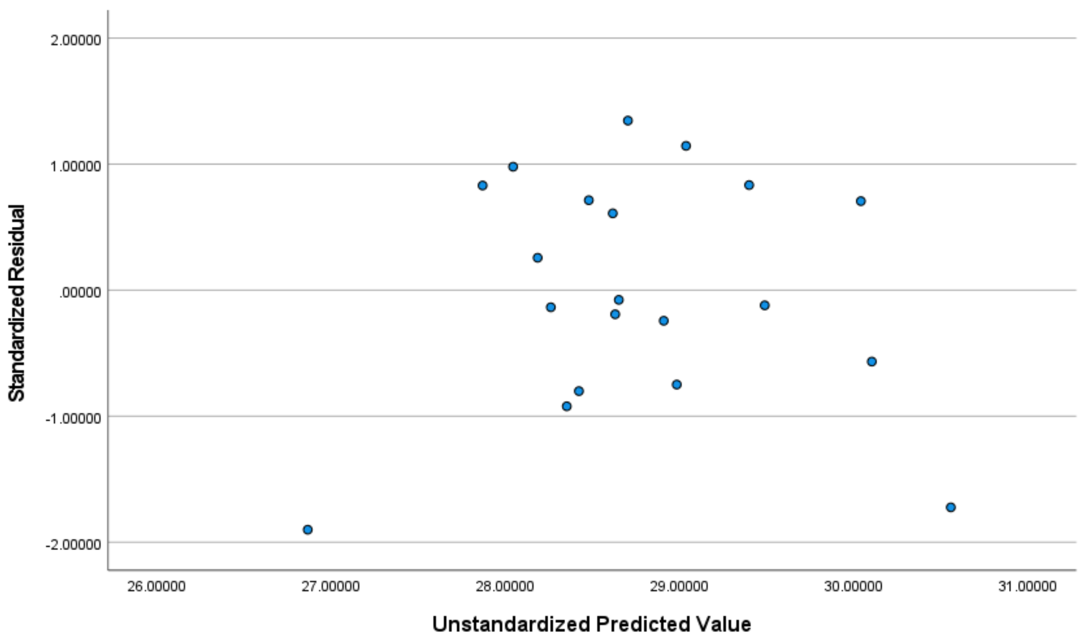
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13.633	3	4.544	162.532	.000[b]
	Residual	.447	16	.028		
	Total	14.080	19			

a. Dependent Variable: Y

b. Predictors: (Constant), Trim, Weight, Fat

Residual Plot:



CONCLUSION:

The dependency of Y was studied using the following:

- **Scatter Plot:** There appears an approximate decreasing line on the plot between Y and Fat which means a moderate relationship. While plotting Y and Weight, the plot looks scattered all over indicating no relationship. Y is strongly dependent on Trim as there is a linear dotted line in the scatter plot.

- **Normality:** Based on the information above, it seems like all the data is normally distributed for all the variables.

- **F-Test:** The critical value was found to be 3.24 from the F distribution table. We reject the null hypothesis (H_0) if F $\geq$ c = 3.24 (F^3_{16} & $\alpha = 5\%$). From the table of ANOVA provided above, the value of F statistic is 162.532. In this case we reject the null hypothesis meaning that there is a strong relationship between the dependent and the independent variables.

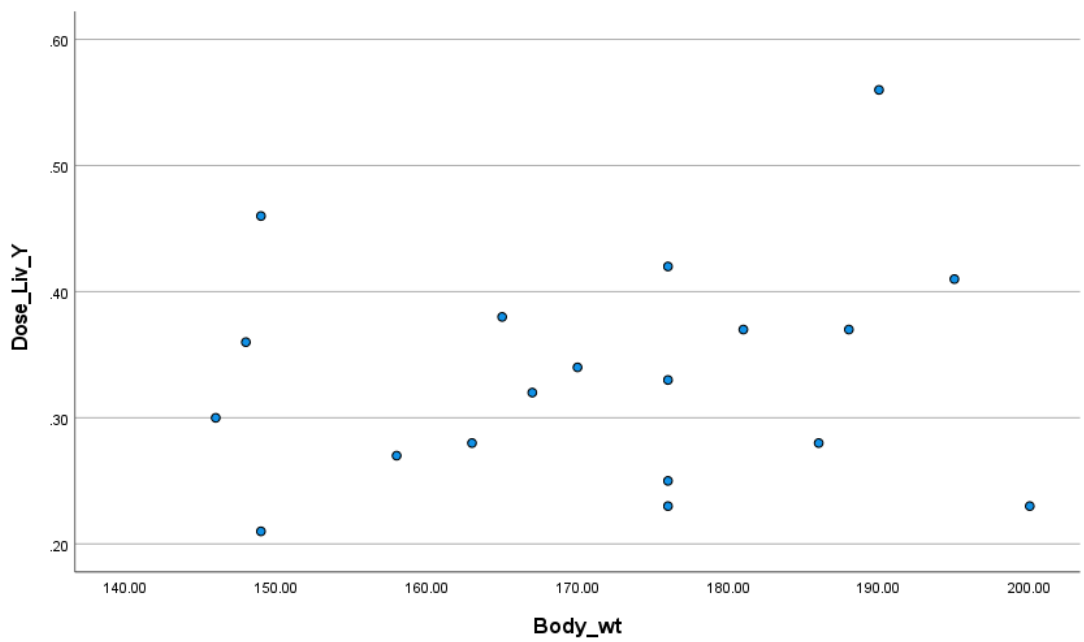
- **T-Test:** The critical value was found to be 1.746 from the t distribution table. We reject the null hypothesis (H_0) if -c $\geq$ T $\geq$ c = 1.746 (df=16 & $\alpha = 5\%$). From the table of Coefficients provided above, the value of t statistic for fat, weight and trim are 0.336, -0.525 and 14.590 respectively. In this case we accept the null hypothesis of fat and weight while we reject the null hypothesis of trim. Thus the conclusion is that fat and weight are useful for determining the value of Y while trim cannot be used for explaining Y. This holds true while considering the value of sig for the three variables (sig>0.05 => accept null hypothesis; sig<0.05 => reject null hypothesis).

- **Residual Plot:** As there are no outliers found (limit +2 and -2) and the residual plot here is an usual random scatter plot, it seems to be a good linear regression model.
- **Adjusted R squared:** The value of adjusted R squared is 0.968 (or) 96.8%. This means that the dependent variable is 96.8% described by the independent variables. This value seems to be very large for me and so I conclude that there is a strong relationship between the independent variables and the dependent variable (Y).

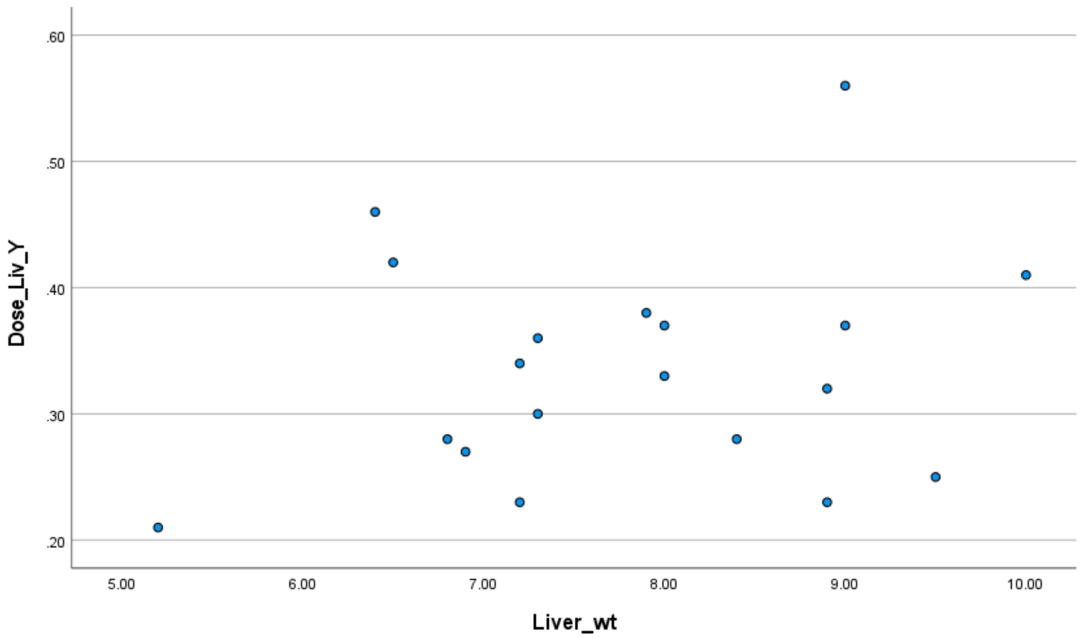
RATS

Scatter Plot:

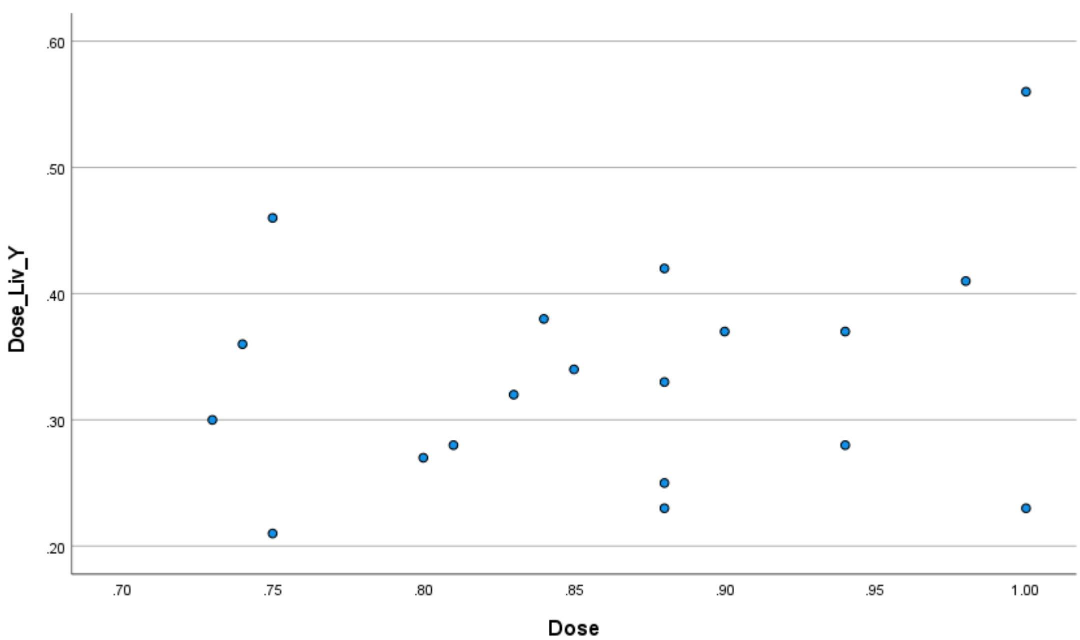
Y - Body Weight



Y - Liver Weight



Y - Dose



Normality Test:

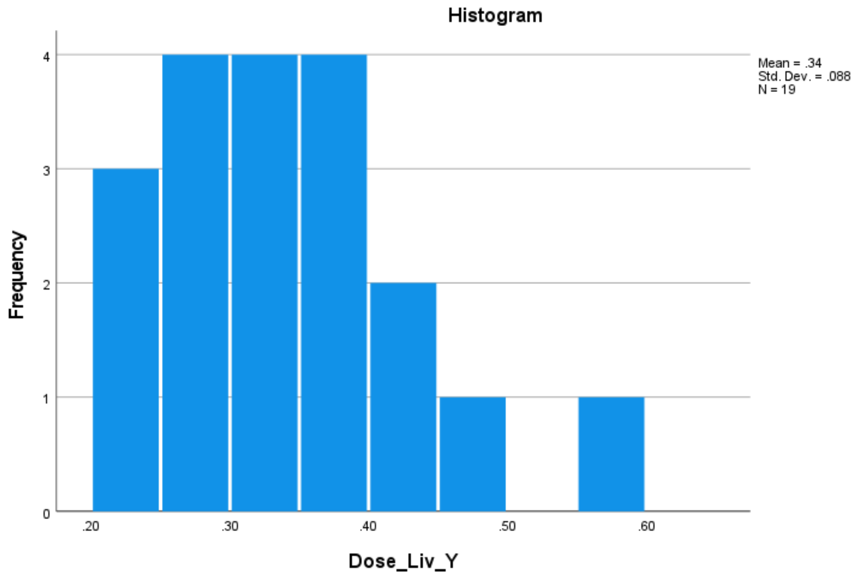
Dose Liver Y

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Dose_Liv_Y	.102	19	.200*	.953	19	.439

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Body Weight

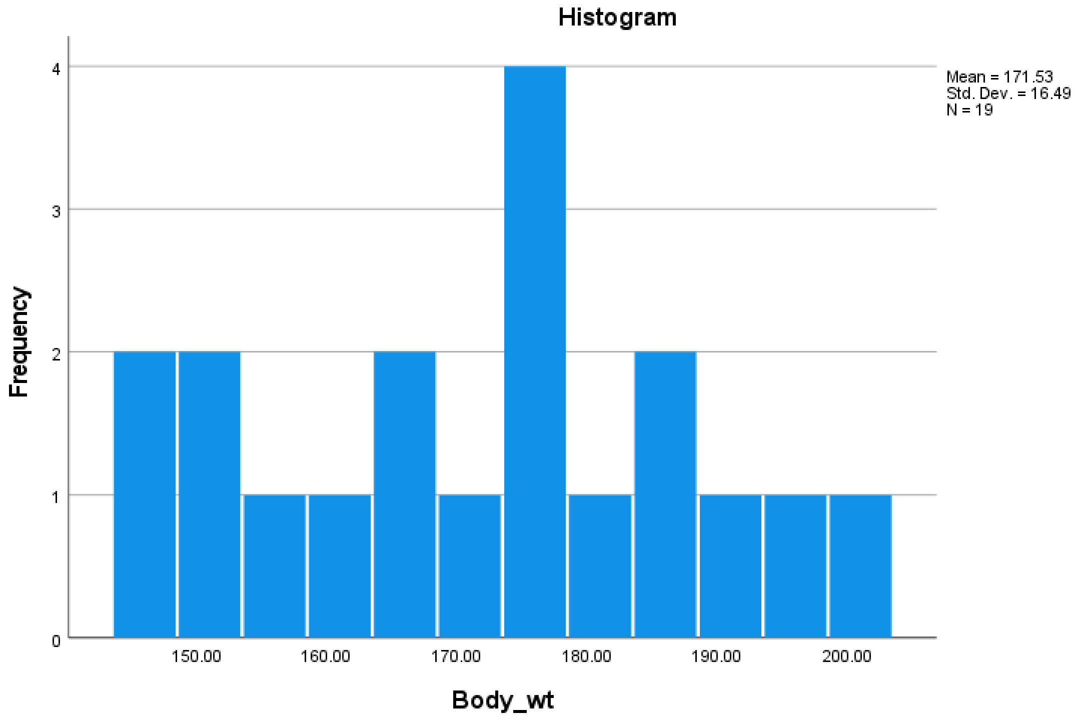
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Body_wt	.133	19	.200*	.954	19	.453

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Liver Weight

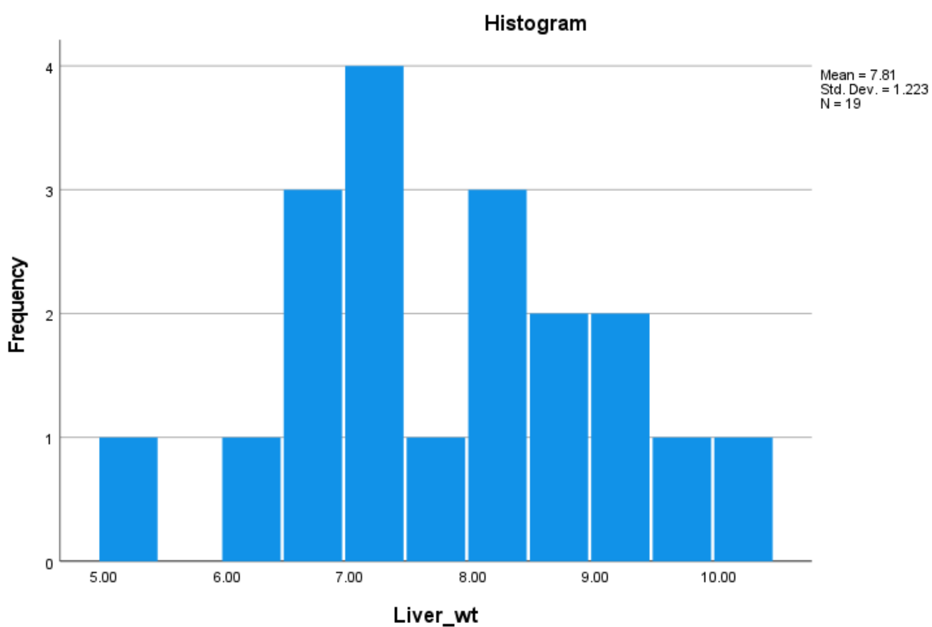


Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Liver_wt	.136	19	.200*	.975	19	.871

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



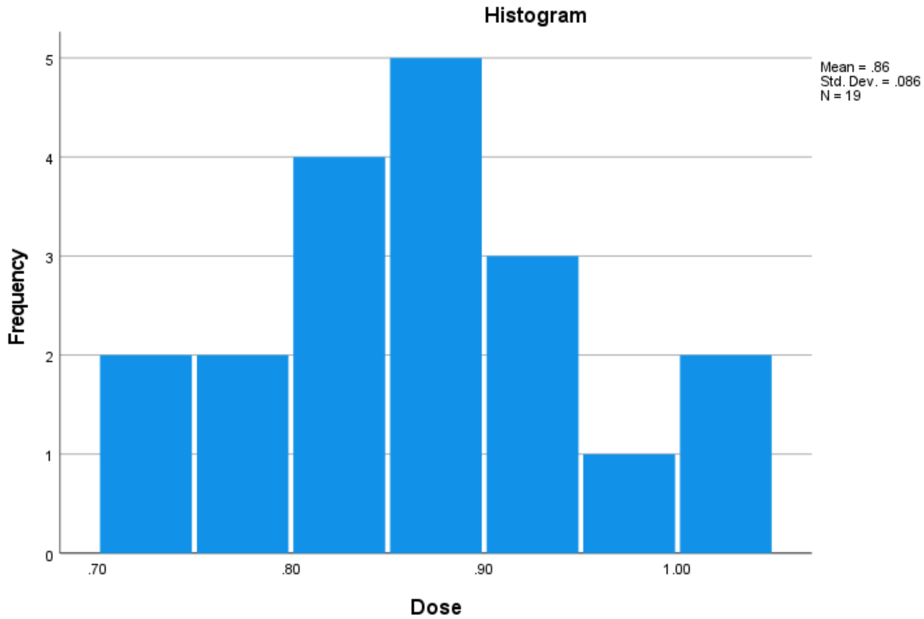
Dose

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Dose	.115	19	.200*	.947	19	.352

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Linear Regression:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.266	.195		1.367	.192
	Body_wt	-.021	.008	-3.960	-2.664	.018
	Liver_wt	.014	.017	.198	.830	.419
	Dose	4.178	1.523	4.052	2.744	.015

a. Dependent Variable: Dose_Liv_Y

Residual Plot:

CONCLUSION:

The dependency of Y was studied using the following:

- **Scatter Plot:** The plots of the dependent variable versus the three independent variables looks scattered all over indicating no relationship.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.603[a]	.364	.237	.07729

a. Predictors: (Constant), Dose, Liver_wt, Body_wt

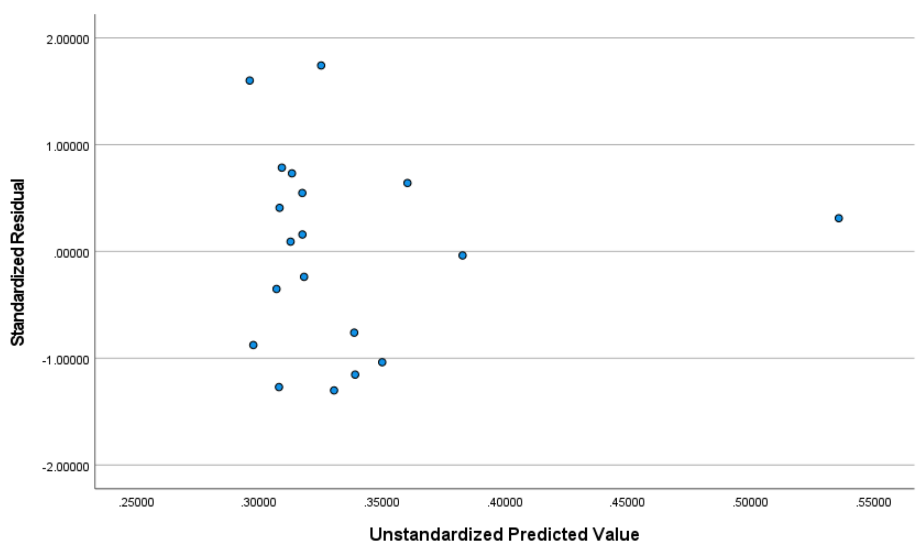
b. Dependent Variable: Dose_Liv_Y

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.051	3	.017	2.860	.072[b]
	Residual	.090	15	.006		
	Total	.141	18			

a. Dependent Variable: Dose_Liv_Y

b. Predictors: (Constant), Dose, Liver_wt, Body_wt



- **Normality:** Based on the information above, it seems like all the data is normally distributed for all the variables.

- **F-Test:** The critical value was found to be 3.29 from the F distribution table. We reject the null hypothesis (H_0) if F $\geq$ c $=$ 3.29 (F^3_{15} & $\alpha = 5\%$). From the table of ANOVA provided above, the value of F statistic is 2.860. In this case we do not reject the null hypothesis meaning that there is a very weak relationship between the dependent and the independent variables.

- **T-Test:** The critical value was found to be 1.753 from the t distribution table. We reject the null hypothesis (H_0) if -c $\geq$ T $\geq$ c $=$ 1.753 (df=15 & $\alpha = 5\%$). From the table of Coefficients provided above, the value of t statistic for body weight, liver weight and dose are -2.664, 0.830 and 2.744 respectively. In this case we reject the null hypothesis of body weight and dose while we accept the null hypothesis of liver weight. Thus the conclusion is that body weight and dose are not useful for determining the value of dose in liver while liver weight can be used for explaining the dose in liver. This holds true while considering the value of sig for the three variables (sig>0.05 => accept null hypothesis; sig<0.05 => reject null hypothesis).

- **Residual Plot:** As there are no outliers found (limit +2 and -2) and the residual plot here is an usual random scatter plot, it seems to be a good linear regression model.

- **Adjusted R squared:** The value of adjusted R squared is 0.237 (or) 23.7%. This means that only 23.7% of the dependent variable is described by the independent variables. This value seems to be very small for me and so I conclude that there is a very weak relationship between the independent variables and the dependent variable (Dose Liver Y).