UNIVERSITY OF TEXAS ARLINGTON
Industrial, Manufacturing, & Systems Engineering Department
College of Engineering

IE5318 - 001
Multiple Linear Regression Project
Date: 2nd May 2022
Topic: Factors affecting the weight of humans

#	Group Members' Name	Student ID
1	Jatin Sharma	1002071716
2	Paradkar, Pankaj	1001851313
3	Togare, Sumit	1001968052

1. Proposal

1.1 Description of the problem and variables

The idea behind this project is to describe what factors affect the weight of humans by selecting a sample population. The sample population selected for this project are the students at the University of Texas at Arlington (UTA). Four factors have been identified for this project, they are defined and the reason for selecting them are mentioned below. Here, X1 to X4 are the predictor variable and Y is the response variable.

X1: Gender – It is defined as the characteristics of men or women. According to the National Institute of Diabetes and Digestive and Kidney Diseases a part of the U.S. Department of Health and Human Services, obesity is more common in woman than men because of where (location) the body stores fat.[1]

X2: Height – It is defined as the measurement of a UTA student from feet to top of his/her head. As a person's height increases, their body volume also increases which results in increase in mass i.e., weight.

X3: Sleeping time in a day – It is defined as the amount of time a UTA student sleeps in a day. According to Amy Shapiro, when a person does not sleep well, he/she will feel tired. Feeling tired make the body need more energy. As a result of this, people tend to eat more and put on weight.[2]

X4: Amount of water consumed in a day – It is defined as the amount of water consumed by a UTA student in a day. Lower water consumptions lead to lower metabolism rate which decreases lipolysis resulting in higher weight says.[2]

Y: Weight – It is defined as the measurement of mass of a UTA student. It is the response to the change for any one or all the above factors.

1.2 Data Collection

The data for the project was collected through a survey. Details of the survey can be found in the appendix section 9.1 of this report. Two locations within UTA were setup, the University Centre, and the Maverick Activity Centre. Students visiting these two locations were approached randomly and was requested to take a survey by scanning a QR code. The survey began by taking the consent of the respondent for providing their personal information. Next, the survey had a series of questions regarding participants gender, height, sleeping time, water consumption and weight.

The data collected from the survey is in the appendix section 9.1.3 of this report. The survey received responses from 91 students. Among those, the tallest student was 195 cm, and the shortest student was 125.27 cm. The heaviest student was 95 kg, and the lightest student was 43 kg. Minimum water consumption and sleep time among the students was found to be 0.5 litres and 1 hour respectively and the maximum for the same was 15.14 litres and 9 hours, respectively. Finally, male students were 55 and female students were 36.

All the entries made by the students who took part in the survey were not measured but it was self-reported. The self-reported entries are assumed to be accurate for the SLR and MLR analysis.

1.3 Matrix scatter plot of the variables and correlation analysis

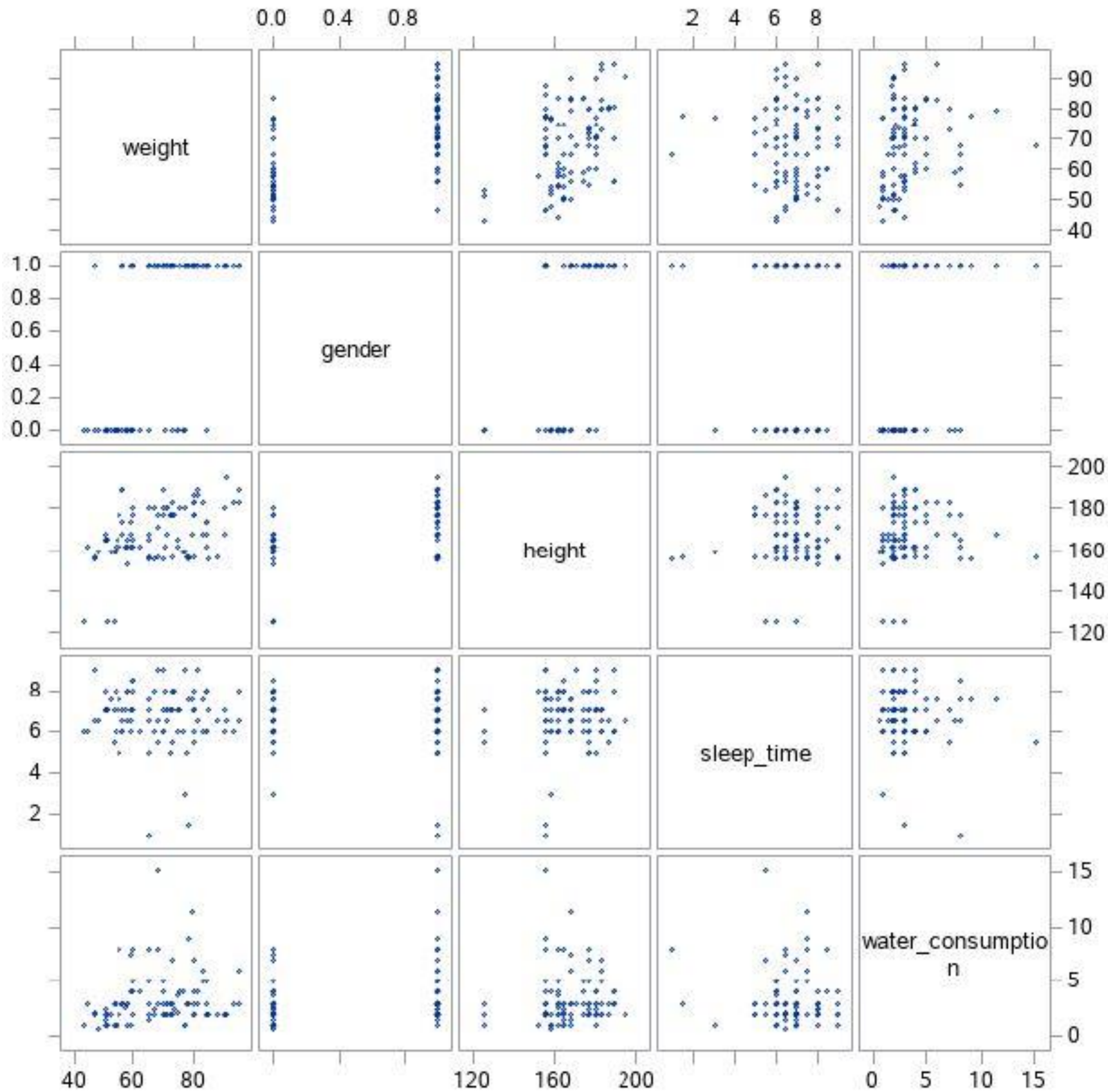

Figure 1: Matrix scatter plot of the variables

Pearson Correlation Coefficients, N = 91				
	Weight	Height	Sleep	Water Consumption
Weight Weight	1.00000	0.43853	-0.05739	0.16655
Height Height	0.43853	1.00000	0.09961	0.03379
Sleep Sleep	-0.05739	0.09961	1.00000	-0.09622
Water Consumption Water Consumption	0.16655	0.03379	-0.09622	1.00000

Table 1: Correlation Analysis

Using Figure 1 and Table 1 together, first let us look at y – x plots. In these plots it is desired to have a linear trend among weight of UTA students and the independent variables. From the plot we can say that the weight vs height is linear in relationship with a positive trend. This is corroborated with the correlation at 0.43853 which can be categorized as medium correlation which is good. Some outliers are identified, during the survey it was observed that some of the students were indeed short (between height 120 to

130 cm) and have appeared as outliers. The outliers in this case are true and not an error entry. Next, weight vs sleep, from this plot we can say that there is no trend and appears as random point cloud shape. Correlation analysis depicts the same, as low correlation at -0. 05739. Finally, from the plot and correlation analysis we can say that weight vs water consumption has little correlation. We can see from the plot that there is a very little upward trend with correlation of 0.16655.

Now that we have deciphered all the y – x plots, next, we need to look at the x – x plots. With x – x plots, it is desire to have no trend among independent variables i.e., zero correlation and have random point cloud shape. From the plot we can say that the height vs sleep has random point cloud shape with outliers in the x – direction. The correlation analysis also depicts the same picture with low correlation at 0.09961. Next, the height vs water consumption also has random point cloud shape corroborated by the correlation analysis with low correlation of 0.03379 and with some outliers in the x – direction. Finally, the sleep vs water consumption also has random point cloud shape with low correlation -0.09622 and some outliers in the x – direction.

1.4 Potential Complications

Regarding potential complications, there appears to be curvilinearity with weight vs water consumptions. From x – x plots and correlation analysis we can say that there is multicollinearity however, it is low which is good because the model will be stable.

2. Preliminary Multiple Linear Regression Model Analysis

2.1 Preliminary Regression Model

Equation 1 below represents the model form for the multiple linear regression with respect to UTA students' weight (Y), which is influenced by their respective gender(x1), height(x2), sleep time(x3) and water consumption(x4). Here Y_i represents the predicted weight, X_{i1} represents the gender, X_{i2} is the height, X_{i3} is the sleep time and X_{i4} represents the water consumption of a student at UTA.

$$Y_i = \beta_o+\beta_1X_{i1}+ \beta_2X_{i2}+ \beta_3X_{i3}+ \beta_4X_{i4} + \varepsilon_i$$

Equation 1: Regression Model Form

Parameter Estimates								
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Type I SS**	**Variance Inflation**
Intercept	Intercept	1	30.34441	15.74436	1.93	0.0572	422002	0
g	gender	1	13.45582	2.60285	5.17	<.0001	5783.55853	1.37548
h	height	1	0.20047	0.09394	2.13	0.0357	434.73761	1.32137
s	sleep	1	-0.71036	0.82550	-0.86	0.3919	87.84001	1.02155
wc	water_consumption	1	0.21778	0.47027	0.46	0.6445	22.98371	1.06932

Table 2: Preliminary model parameter estimates

The Table 2 is obtained using SAS, parameter estimates are used to build the preliminary regression model which is stated below:

$$Y_i = 30.34441 + 13.45582X_{i1} + 0.20047X_{i2} - 0.71036X_{i3} + 0.21778X_{i4} + \varepsilon_i$$

Equation 2: Preliminary Regression Model

2.2 Check for model assumptions

2.2.1 Check for constant variance assumption

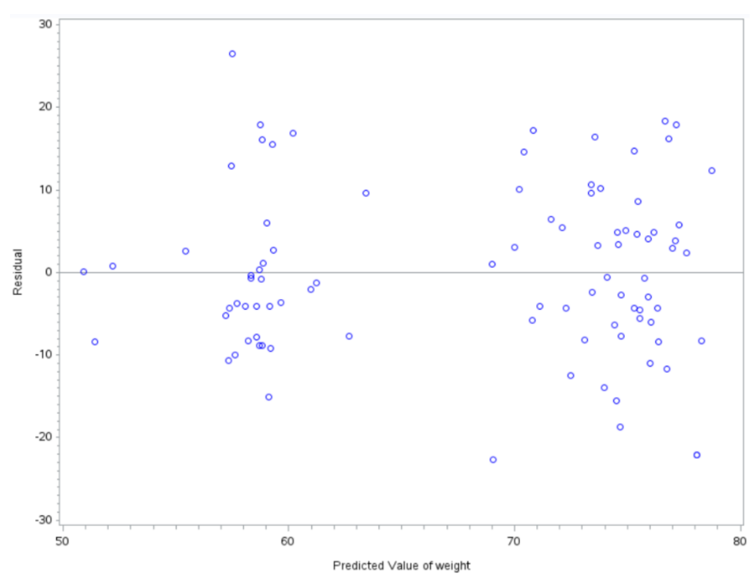

Figure 2: Residual vs ŷ plot

It is assumed that the residuals of UTA students' model have constant variance. Figure 2 is used to check for constant variance assumption using the residual vs ŷ plot. Here in Figure 2, we can see that the residual plot has random point cloud shape. With no funnel shape in the residual plot, we can conclude that the UTA students' model satisfies constant variance assumption.

2.2.2 Check for model form

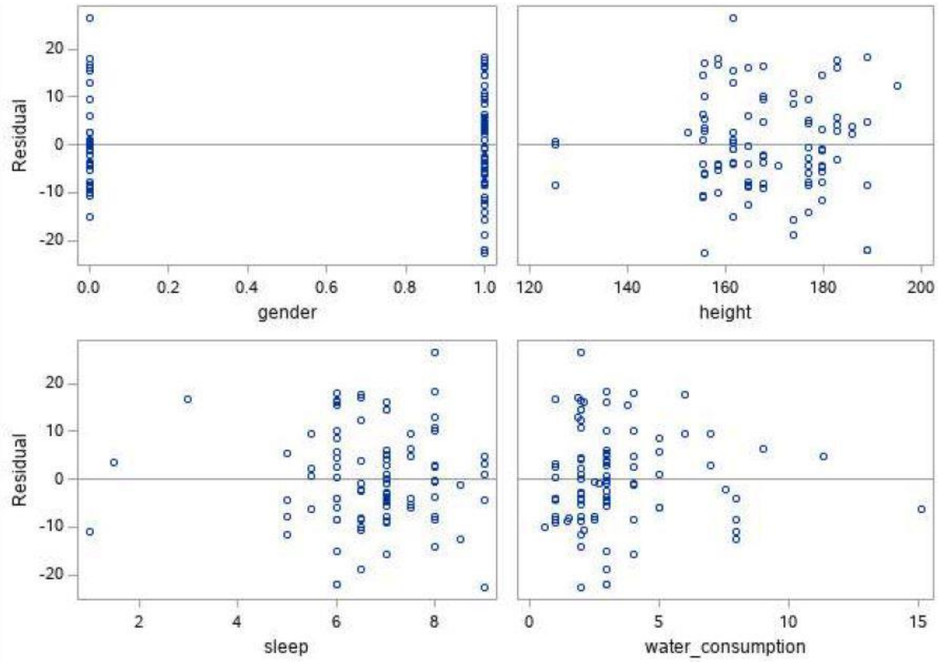

Figure 3: Residual vs Predictor Variable

It is assumed that the current multiple linear regression model form of UTA students is adequate. Figure 3 is used to check this model form and by reviewing all the four plots we can conclude that there is no curvature in any of the residual vs predictor plots. Therefore, we can conclude that the multiple linear regression model form is adequate.

2.2.3 Check for normality assumption

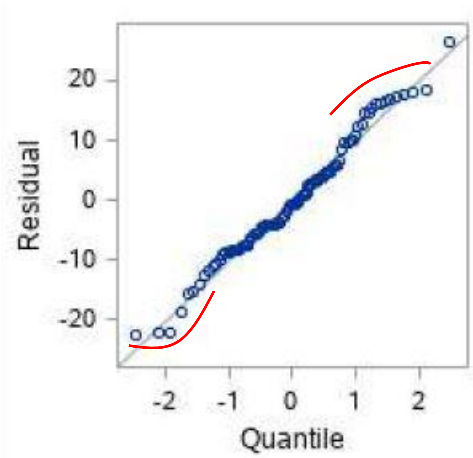

Figure 4: NPP Plot

It is assumed that the UTA students' model is normally distributed. To validate this assumption, NPP plot is constructed using SAS. From Figure 4 we can say that the NPP plot is has a slight shorter right tail and shorter left tail than the normal distribution but, overall straight. Hence, we can conclude that normality is close enough, although not perfect. Therefore, normality is not satisfied.

2.2.4 Uncorrelation assumption

Regarding correlation assumption, the weight of UTA students is not collected in time order hence, time plots cannot be generated. Therefore, correlation assumption cannot be checked. Time series in not meaningful.

2.3 Model Diagnostics

2.3.1 Outliers, Leverage, and Influence

This section of the report diagnoses the UTA students' model and checks for outliers - its leverage and influence on the model. The outliers are looked out for in both x and y direction. To check the outliers in the y-direction, a comparison between t-distribution and R Student is done. If any observations in the R Student is greater than the t-distribution, then the observation is flagged as an outlier in the y-direction. Similarly for the outliers in the x-direction, a comparison between h_{ii} and $\frac{2p}{n}$ is done. If any observations in the h_{ii} is greater than $\frac{2p}{n}$, then the observation is flagged as an outlier in the x-direction. If an observation is identified as an outlier, then its influence on the fitted values is considered on the regression by using DFFITS and comparing it with cutoff value $2\sqrt{\frac{p}{n}}$ since n is large (n >=30). Next, the outliers influence on the individual LSEs is considered by using DFBETAS and comparing it with cutoff values $\frac{2}{\sqrt{n}}$ since n is large. Finally, Cooks Distance is also used to identify the influence of the outliers on the model. It is similar to DFBETAS however, Cooks Distance is looking for combined impact of all LSCs in the UTA students' model. If an observation D_i is greater than the cutoff value F (0.5; p; n-p), that observation is labeled as influential. The necessary cut-off values are calculated using SAS with 99% CI and shown below in Table 3.

Obs	tinvtres	finv50
1	3.37931	0.87717

Table 3: Cutoff values

Table 4 below is generated using SAS and cutoff values from Table 3 is compared to identify outliers and its influence on the model. Cuttoff for h_{ii}, DFFITS and DFBETAS is found to be $\frac{2p}{n} = \frac{2*5}{91} = 0.1098$, $2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{5}{91}} = 0.4688$ and $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{91}} = 0.2096$ respectively.

Table 4: Output Statistics

Output Statistics											
Obs	**Residual**	**Cooks Di**	**RStudent**	**Hat Dlag H**	**Cov Ratio**	**DFFITS**	**DFBETAS**				
							Intercept	**Gender**	**Height**	**Sleep**	**Water Consumption**
5	-8.4128	1.22E-08	-0.8739	0.1376	1.1756	0.3491	-0.3195	0.0549	0.2971	0.0296	0.0994
17	3.3890	5.35E-11	0.3658	0.2073	1.3271	0.1871	0.1066	0.0583	-0.0474	-0.1656	-0.0347
19	16.9066	4.22E-09	1.7683	0.1259	1.0125	0.6711	0.2321	-0.1817	0.0217	-0.5733	-0.1970
22	0.7964	2.03E-09	0.0820	0.1301	1.2183	0.0317	0.0293	0.0031	-0.0268	-0.0056	-0.0016
35	0.1089	7.20E-10	0.0112	0.1320	1.2215	0.0044	0.0037	0.0006	-0.0039	0.0006	-0.0006
47	-6.3750	3.76E-09	-0.7311	0.2944	1.4562	-0.4722	-0.0660	-0.0216	0.0941	0.0304	-0.4313
48	4.8269	2.23E-11	0.5020	0.1448	1.2215	0.2065	-0.0180	0.0100	-0.0165	0.0530	0.1878
72	-10.9940	1.28E-09	-1.2326	0.2532	1.2994	-0.7178	-0.3286	0. 1269	0.1379	0.6025	-0.1982
73	9.5991	4.69E-10	0.9788	0.1030	1.1176	0.3317	0.1364	-0,2406	0.1824	-0.1007	0.2053
88	6.4056	1.2E-05	0.6511	0.1029	1.1528	0.2205	0.0609	0,0666	-0.1025	0.0687	0.1585

Every observation in the Cooks Distance does not exceed 0.88240. Cooks distance does not detect influence. Similarly, every observation in R Student does not exceed 3.37931. Hence, we can conclude that there are no y-outliers. Observations 5, 17, 19, 22, 35, 47, 48, 72, 73 and 88 are greater than 0.1098 therefore, we can conclude that these observations are outliers in the x-direction. Next, the corresponding |DFFITS| is compared with $2\sqrt{\frac{p}{n}} = 0.4688$ and observations 19, 47 and 72 are greater than 0.4688 therefore, we can conclude that these observations influence the fitted value ($\hat{y}_i$). Finally, the corresponding |DFBETAS| is compared with $\frac{2}{\sqrt{n}} = 0.2096$ and observations 19, 47 and 72 are greater than 0.2096 therefore, we can conclude that these observations influence the LSEs.

Now that we have identified influential outliers, lets discuss them interns of UTA students' application. Observation 19 and 72 are the outliers for sleep time predictor variable. The sleep time recorded for observations 19 and 72 are 3 hours and 1 hour respectively. This is quite bizarre because it is impossible for someone to sleep for just 3 hours or 1 hour every day. Hence, we can conclude that the respondent has provided false information in the survey and therefore, observation 19 and 47 are outliers. Next, observation 47 is outlier for water consumption. The water consumption recorded for this observation is 15.14 liters. It is impossible for a person to consume this much quantity of water every day and therefore we can conclude that observation 47 is an outlier.

2.3.2 Variance Inflation Factor (VIF)

Parameter Estimates								
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Type I SS**	**Variance Inflation**
Intercept	Intercept	1	30.34441	15.74436	1.93	0.0572	422002	0
g	gender	1	13.45582	2.60285	5.17	<.0001	5783.55853	1.37548
h	height	1	0.20047	0.09394	2.13	0.0357	434.73761	1.32137
s	sleep	1	-0.71036	0.82550	-0.86	0.3919	87.84001	1.02155
wc	water_consumption	1	0.21778	0.47027	0.46	0.6445	22.98371	1.06932

Table 5: Parameter Estimates

Moving on with the model diagnostics, next, we need to assess how sever multicollinearity is. To assess this, VIF is calculated using SAS and the output is represented in Table 5. This is calculated by regressing each predictor on the remaining predictors. From Table 5 we firstly have, $(VIF)_1$ = 1.37548 where gender is regressed on height, sleep time and water consumption. Secondly, we have, $(VIF)_2$ = 1.32137 where height is regressed on gender, sleep time and water consumption. Next, we have, $(VIF)_3$ = 1.02155 where sleep time is regressed on gender, height, and water consumption. Finally, we have, $(VIF)_4$ = 1.06932 where water consumption is regressed on gender, height, and sleep time.

Next, we need to find $(VIF)_k$ value by using the formula $(\overline{VIF}) = \frac{\sum_{k=1}^{p-1}(VIF)_1}{p-1} = \frac{1.37548+1.32137+1.02155+1.06932}{4-1} = 1.5$

Since $(VIF)_k$ = 1.5 and max (VIF) = 1.37548 which is little more than 1.0, we can conclude that serious multicollinearity is not a problem.

2.4 Model Assumption Tests

2.4.1 Modified – Levene Test

Modified-Levene test is test for non-constant variance. For this test, the data needs to be divided into two groups. The two groups are formed by first calculating the ŷ for each observation and the median of ŷ was found to be 72.1 kgs. Next, observations with ŷ < 72.1 kg was formed as group 1 with 45 observations and observations with ŷ > 72.1 kg was formed as group 2 with 46 observations.

Obs	group	mede
1	1	-0.76432
2	2	-0.64482

Table 6: $\tilde{e}1$ and $\tilde{e}2$ Values

group	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		45	7.5539	6.5740	0.9800	0	27.2833
2		46	8.4551	5.7903	0.8537	0.1040	21.4297
Diff (1-2)	Pooled		-0.9013	6.1902	1.2979		
Diff (1-2)	Satterthwaite		-0.9013		1.2997		

Table 7: T Test Procedure

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	89	-0.69	0.4892
Satterthwaite	Unequal	87.085	-0.69	0.4899

Table 8: T Test Procedure

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	44	45	1.29	0.3996

Table 9: T Test Procedure

From Table 6 we get $\tilde{e}_1$ = -76432 and $\tilde{e}_2$ = -0.64482
From Table 7, and 9 we get σ_{12} = 6.5740, σ_{22} = 5.7903, F' = 1.29, DF = (44, 45), Prob > F' = 0.3996

F - Test at $\alpha = 0.05$
H_0: σ1 = σ2 vs H_1: σ1 ≠ σ2
As 0.3996 > 0.05 we have P > α
Fail to Reject H_o. This is a weak conclusion
Therefore, the variance could be equal.

T Test: T-Test has two forms of the two sample T – Test and the F – Test is used to choose which form. In F-Test, since we concluded variance is equal, we use equal form to do two sample T-Test.
H_0: Means of d_{i1} and d_{i2} populations are equal vs H_1: Means of d_{i1} and d_{i2} populations are not equal
As variance may be equal, from Table 8 we have Prob>|T| as 0.4892
P = 0.4892 and α = 0.05
As 0.4892 > 0.05 we have P > α
Fail to Reject H_o and this is a weak conclusion,
Modifies-Levene test does not detect non-constant variance.
Therefore, variance is Constant.

The result of the Modified-Levene test is that the variance is constant. Figure 2 corroborates the test since the residual plot does not have a funnel shape. From Modified-Levene test and Figure 2 we can concludes that the constant variance assumption of the UTA students' model is satisfied.

2.4.2 Normality Test

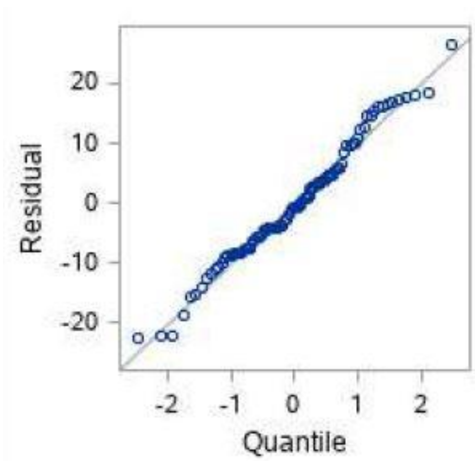

Figure 5: NPP

Pearson Correlation Coefficients, N = 91 Prob > |r| under H0: Rho=0		
	e	enrm
e Residual	1.00000	0.99227 <.0001
enrm 'Normal Scores'	0.99227 <.0001	1.00000

Table 10: Sample Correlation

H_0: Normality is OK vs H_1: Normality is NOT OK
Cutoff c (α, n) = c (0.1, 91) = 0.988
From Table 10 we get sample correlation $\hat{\rho}$ = 0.99227
We have, 0.99227 > 0.988 i.e., $\hat{\rho} > c(\alpha, n)$
We fail to reject H_o, this is a weak conclusion and conclude normality is OK, which means the test does not detect nonnormality.

The normality test does not detect nonnormality. By looking at Figure 5 we can say that the NPP plot is has a slight shorter right tail and shorter left tail than the normal distribution but, overall straight. Hence, we can conclude that normality is close enough, although not perfect. Therefore, normality is not satisfied. By this, we can conclude that the variation of observed value of UTA students' weight to the actual value of the weight along the regression line is close to normal distribution, although not perfect.

2.4.3 Breusch-Pegan Test

From Figure 5 and normality test in section 2.4.2, we have concluded that normality assumption is not satisfied, because of this we have conducted the test for constant variance using Modified – Levene Test and not Breusch - Pagan.

2.5 Model assumptions satisfied preliminary model

From section 2.2 through 2.4 we have verified model assumptions and concluded that the UTA students' model has constant variance, the model form is adequate, and normality is not satisfied (satisfying normality is not necessary, if it did, it is a bonus). We have also performed diagnostics – variance inflation and checked for outliers – its leverage and influences. Now, this section of the report presents the preliminary model that satisfied the model assumptions, and its regression output is discussed. Equation 3 below is the preliminary model that satisfied the model assumption.

$$Y_i = 30.34441 + 13.45582X_{i1} + 0.20047X_{i2} - 0.71036X_{i3} + 0.21778X_{i4} + \varepsilon_i$$

Equation 3: Model Assumption Satisfied Preliminary Regression Model

Table 12 below represents the ANOVA table for the presented preliminary regression model.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	6329.11986	1582.27997	14.76	<.0001
Error	86	9216.48469	107.16843		
Corrected Total	90	15546			

Table 11: ANOVA: Preliminary Regression Model

To test the significance of the preliminary regression model we can use only F-Test. The significance is tested at $\alpha = 0.1$ below:
H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots\ldots\ldots\ldots\ldots \beta_{p-1} = 0$ which is a reduced model $Y_i = \beta_0 + \varepsilon_i$
H_1: at least one β_k is not zero which is a full model $Y = x\beta + \varepsilon$
From Table 11, we have F-Statistic: F^*=14.76 and from F-Distribution we have $F(1 - \alpha; p\text{-}1; n\text{-}p) = F(1 - 0.1; 4; 86) = 3.55$
Since $F^* = 14.76 > F(1 - \alpha; p\text{-}1; n\text{-}p) = 3.55$ we can conclude that preliminary regression model is significant. This is a strong conclusion. From this we can infer that the preliminary regression model is significant because any one of the predictor variables – gender, height, sleep time and water consumption is important. However, we cannot conclude exactly which predictor is most important with this test.

From Table 12 we have the sum of square regression (SSR) equal to 6329.11986 which is the deviation of predicted weights ($\hat{y}$) of UTA students from their mean weight ($\overline{y}$) of all 91 students which are explained by the predictor variables - gender, height, sleep time and water consumption. The sum of square error (SSE) is equal to 9216.48469 and it is the deviation of the observed weight (y) from the mean weight ($\overline{y}$) of all 91 students which is not explained by the predictor variables gender, height, sleep time and water consumption and considered as errors.

Now we can check what percentage of total variation from mean weight of UTA students is explained by this model by calculating coefficient of determination which is measured using the equation $R_2 = \frac{SSR}{SSTO}$, R_2 is found out to be 0.4071 which implies that only 40.71% of variation in weight of the UTA students is explained by this model using predictor variables - gender, height, sleep time and water consumption.

3. Exploration of Interaction Terms

This section of the report is dedicated to exploring and finding possible interaction terms to add to the preliminary regression model that satisfied the model assumptions. This is important because, adding interactions helps understand the relationship between UTA students' gender, height, sleep time and water consumption and this explains how a combination of two or more of these independent variables together influence the weight of UTA students.

3.1 Partial Regression Plots

In exploring the interaction terms, the first step is to identify trend among the partial regression plots. The desired trend is linear. Below are the 6 partial regression plots discussed, which are obtained using SAS.

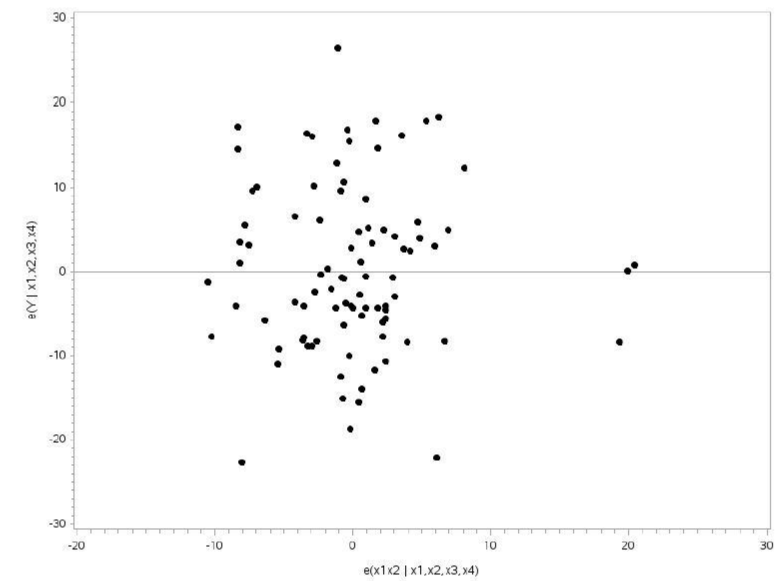

Figure 6: Partial Regression Plot e e(Y|x1,x2x3x4) vs e(x1x2|x1,x2x3x4)

By examining Figure 6, we can conclude that there is no linear trend in the plot and therefore, it is not useful to combine and add gender and height to the preliminary regression model that satisfied the model assumptions based on partial regression plots.

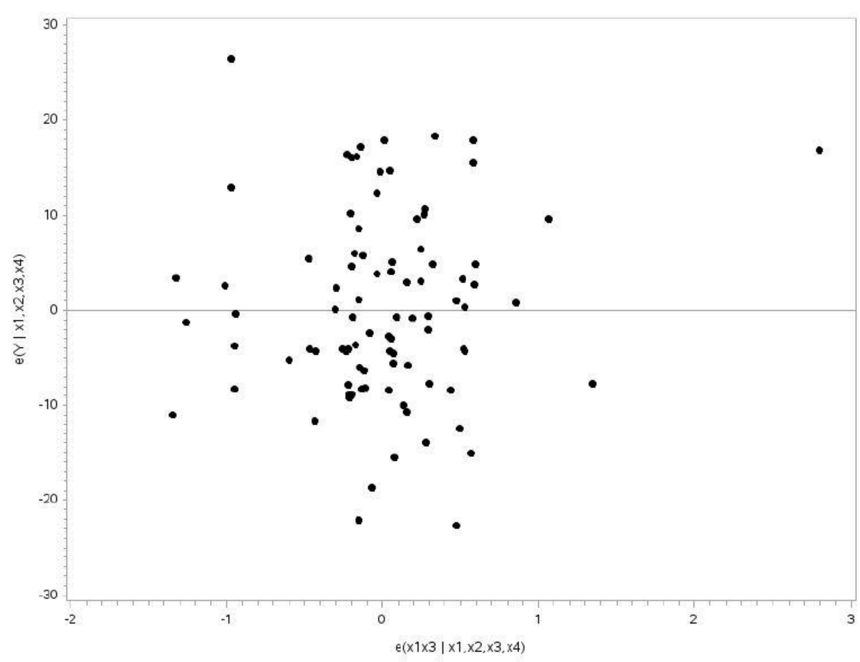

Figure 7: Partial Regression Plot e(Y|x1,x2x3x4) vs e(x1x3|x1,x2x3x4)

By looking at Figure 7 we can conclude that there is no linear upward pattern, and therefore not useful to combine and add gender and sleep time to the preliminary regression model that satisfied the model assumptions based on the partial regression plot.

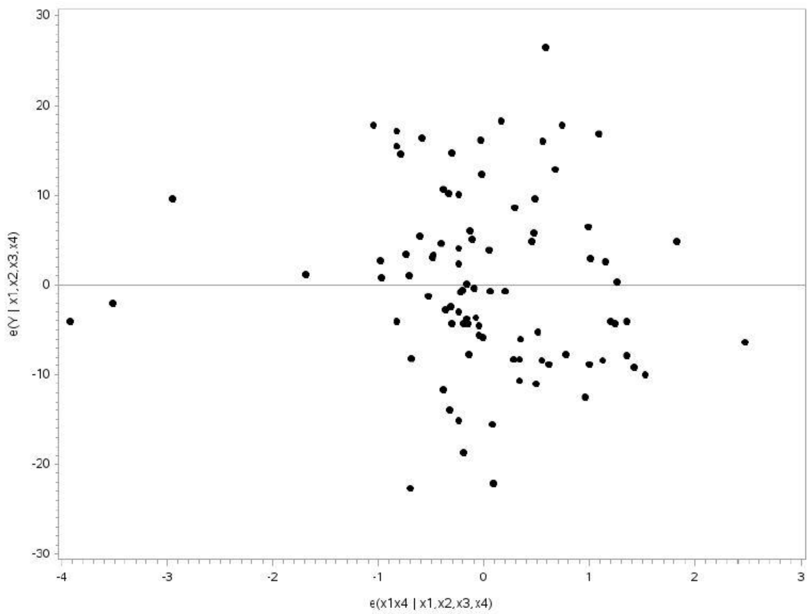

Figure 8: Partial Residual Plot e(Y|x1,x2x3x4) vs e(x1x4|x1,x2x3x4)

In Figure 8 clearly the points are scattered. Hence, we can conclude that combining and adding gender and water consumption to the preliminary regression model that satisfied the model assumptions will not be useful.

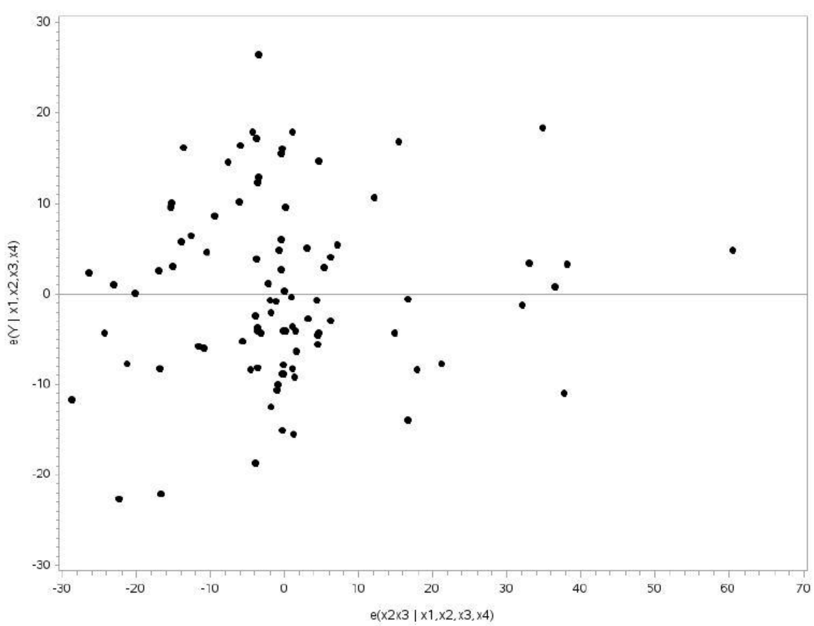

Figure 9: Partial Regression Plot e(Y|x1,x2x3x4) vs e(x2x3|x1,x2x3x4)

By examining Figure 9, we can conclude that there is linear upward trend in the plot, therefore, it is useful to combine and add height and sleep time to the preliminary regression model that satisfied the model assumptions based on partial regression plots.

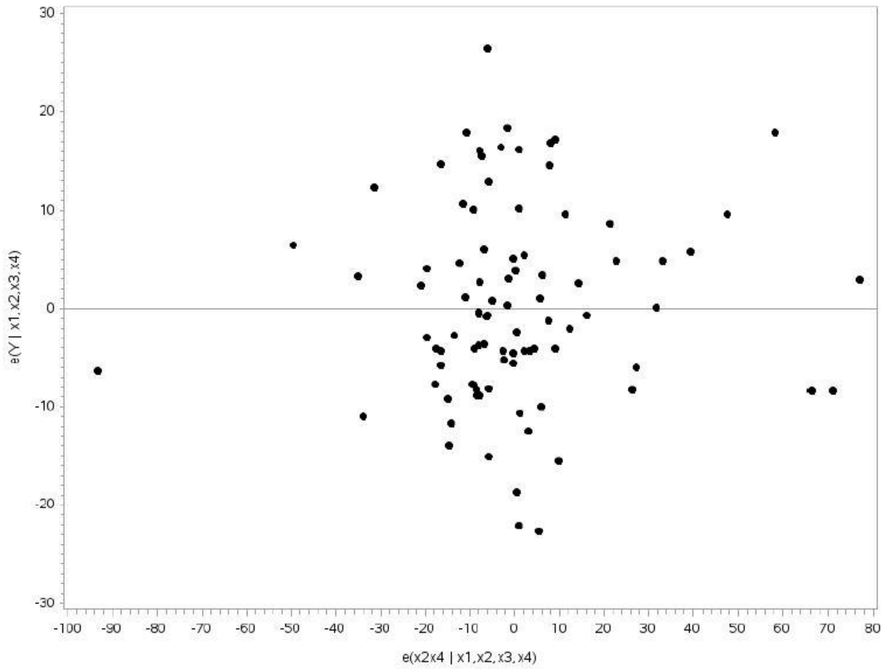

Figure 10: Partial Regression Plot e(Y|x1,x2x3x4) vs e(x2x4|x1,x2x3x4)

Figure 10 illustrates a dense cluster in the middle and shows no linear trend hence, combining and adding height and water consumption to the preliminary regression model that satisfied the model assumptions will not be useful.

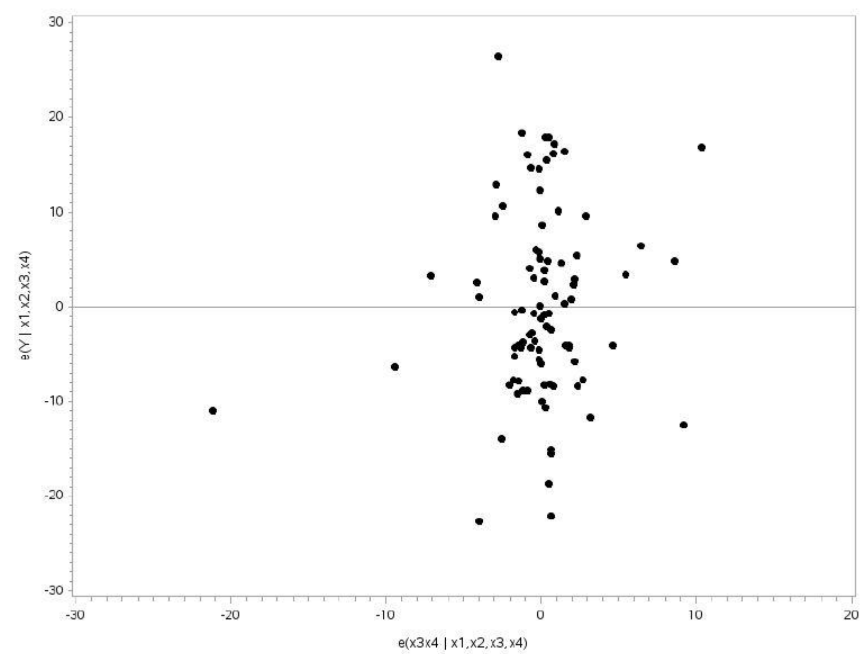

Figure 11: Partial Regression Plot e(Y|x1,x2x3x4) vs e(x3x4|x1,x2x3x4)

Finally, Figure 11 also shows a dense cluster in the middle and shows no linear trend. It will not be worth combining and adding sleep time and water consumption to the preliminary regression model that satisfied the model assumptions.

3.2 Check for correlation

Now that we have decide to bring in interaction terms into the preliminary regression model that satisfied the model assumptions, this might cause multicollinearity problems. Therefore, the next step in exploring interactions is to check correlation between the interaction terms and the original predictors.

Pearson Correlation Coefficients, N = 91											
	weight	gender	height	sleep_time	water_consumption	x1x2	x1x3	x1x4	x2x3	x2x4	x3x4
weight weight	1.00000	0.60995	0.43853	-0.05739	0.16655	0.61807	0.56766	0.38791	0.11322	0.20708	0.15834
gender gender	0.60995	1.00000	0.47816	-0.00325	0.22440	0.99442	0.94331	0.67384	0.17354	0.26290	0.20460
height height	0.43853	0.47816	1.00000	0.09961	0.03379	0.54686	0.47380	0.19863	0.44864	0.13140	0.07453
sleep_time sleep_time	-0.05739	-0.00325	0.09961	1.00000	-0.09622	0.00608	0.28606	-0.09275	0.93178	-0.08139	0.23395
water_consumption water_consumption	0.16655	0.22440	0.03379	-0.09622	1.00000	0.20099	0.17202	0.77874	-0.07141	0.99344	0.91979
x1x2	0.61807	0.99442	0.54686	0.00608	0.20099	1.00000	0.94159	0.65080	0.20691	0.24783	0.18814
x1x3	0.56766	0.94331	0.47380	0.28606	0.17202	0.94159	1.00000	0.60113	0.43158	0.21314	0.26206
x1x4	0.38791	0.67384	0.19863	-0.09275	0.77874	0.65080	0.60113	1.00000	-0.00680	0.78686	0.70180
x2x3	0.11322	0.17354	0.44864	0.93178	-0.07141	0.20691	0.43158	-0.00680	1.00000	-0.02332	0.23618
x2x4	0.20708	0.26290	0.13140	-0.08139	0.99344	0.24783	0.21314	0.78686	-0.02332	1.00000	0.92333
x3x4	0.15834	0.20460	0.07453	0.23395	0.91979	0.18814	0.26206	0.70180	0.23618	0.92333	1.00000

Table 12: Correlation matrix for interaction before standardization

From Table 12, by paying close attention to the correlation between the interaction terms and the original predictors we can say that there are quite a few interaction terms with high multicollinearity (>0.7) which are highlighted. This indicates that there will be a serious multicollinearity issue. Therefore, these interactions cannot be added into the model as it is. To address this issue, the final step in exploration of the interaction terms is standardization. SAS is used to generate standardized interactions and Table 13 is generated to check for high multicollinearity.

Pearson Correlation Coefficients, N = 91											
	weight	gender	height	sleep_time	water_consumption	stdx1x2	stdx1x3	stdx1x4	stdx2x3	stdx2x4	stdx3x4
weight weight	1.00000	0.60995	0.43853	-0.05739	0.16655	-0.17102	0.03004	-0.16400	0.05956	-0.17993	0.03370
gender gender	0.60995	1.00000	0.47816	-0.00325	0.22440	-0.22538	0.00149	-0.10516	-0.00242	-0.29211	-0.08009
height height	0.43853	0.47816	1.00000	0.09961	0.03379	-0.12741	-0.00247	-0.26939	-0.18552	-0.23640	0.05688
sleep_time sleep_time	-0.05739	-0.00325	0.09961	1.00000	-0.09622	-0.00253	0.33957	-0.10375	-0.40277	0.07989	0.23001
water_consumption water_consumption	0.16655	0.22440	0.03379	-0.09622	1.00000	-0.27094	-0.10163	0.30716	0.07076	-0.43373	-0.22537
stdx1x2	-0.17102	-0.22538	-0.12741	-0.00253	-0.27094	1.00000	0.12122	0.03791	0.12054	0.39257	0.15181
stdx1x3	0.03004	0.00149	-0.00247	0.33957	-0.10163	0.12122	1.00000	-0.06505	-0.07748	0.14897	0.40659
stdx1x4	-0.16400	-0.10516	-0.26939	-0.10375	0.30716	0.03791	-0.06505	1.00000	0.10871	-0.07158	-0.12969
stdx2x3	0.05956	-0.00242	-0.18552	-0.40277	0.07076	0.12054	-0.07748	0.10871	1.00000	-0.06581	-0.27636
stdx2x4	-0.17993	-0.29211	-0.23640	0.07989	-0.43373	0.39257	0.14897	-0.07158	-0.06581	1.00000	0.34882
stdx3x4	0.03370	-0.08009	0.05688	0.23001	-0.22537	0.15181	0.40659	-0.12969	-0.27636	0.34882	1.00000

Table 13: Correlation matrix for interaction after standardization

After standardization of the interaction terms now we take a look between the standardized interaction terms and the original predictors in Table 13. Here, we can now see that there is no correlation that are high. Therefore, we can now add standardized form of interaction into the preliminary regression model that satisfied the model assumptions, and this will not create a serious multicollinearity problem by having high VIF (>5).

Since we had decided to bring in x2x3 we can now bring in the standardized form of these interaction into the preliminary regression model that satisfied the model assumptions.

4. Model Search

This section of the report discusses how the required two best models are obtained using Backward Deletion, Best Subsets Selection and Stepwise Regression model search methods. Here the full model is given in equation 4 below:

$$Y_i = 30.34441 + 13.45582X_{i1} + 0.20047X_{i2} - 0.71036X_{i3} + 0.21778X_{i4} + stdx2x3 + \varepsilon_i$$

Equation 4: Full Model

Where X_{i1} is gender, X_{i2} is height, X_{i3} sleep time, X_{i4} water consumption and stdx2x3 std form of height times sleep time of UTA students.

4.1 Best Subset Selection

This is a model with all possible variant of regression. It begins with the full model. For every possible subset of predictor variables, it regresses Y on subset. Once we have all the possible regressions, we need to identify possible best models using the criterions. The criterions are:
1. High R^2 and low SSE.
2. High R_a^2 and low MSE.
3. Generally C_p is required to be close to p, where p is parameters however, low C_p is desired.
4. Finally AIC and SBC are desired to be low.

Using SAS, we now have the different size model in Table 14 for the UTA students. We now need to identify the possible best models using the criterions.

Number in model	Adjusted R-Square	R-Square	C(p)	AIC	SBC	Variables in Model	Model Name	Remark
1	0.365	0.372	3.8723	429.4612	434.4829	gender	A	-
1	0.1832	0.1923	29.8812	452.3662	457.3879	height		Not selected because, C(p) is too large.
Number in model	**Adjusted R-Square**	**R-Square**	**C(p)**	**AIC**	**SBC**	**Variables in Model**		
2	0.3864	0.4	1.8254	427.3156	434.8482	gender height	B	-
2	0.3616	0.3758	5.3331	430.9197	438.4523	gender stdx2x3	C	-
Number in model	**Adjusted R-Square**	**R-Square**	**C(p)**	**AIC**	**SBC**	**Variables in Model**		
3	0.3893	0.4097	2.4262	427.8371	437.8806	gender height stdx2x3	D	-

3	0.3852	0.4057	3.0078	428.4546	438.498	gender height sleep time	E	-
Number in model	**Adjusted R-Square**	**R-Square**	**C(p)**	**AIC**	**SBC**	**Variables in Model**		
4	0.3839	0.4113	4.1946	429.5901	442.1444	gender height sleep time stdx2x3	-	Not selected because, big model.
4	0.3838	0.4112	4.2005	429.5965	442.1508	gender height water consumption stdx2x3	-	Not selected because, big model.
Number in model	**Adjusted R-Square**	**R-Square**	**C(p)**	**AIC**	**SBC**	**Variables in Model**		
5	0.3781	0.4126	6	431.382	446.4472	gender height sleep time water consumption stdx2x3	-	Not selected because, big model.
5	0.3781	0.4126	6	431.382	446.4472	gender height sleep time water consumption stdx2x3	-	Not selected because, big model.

Table 14 Best Subset Regression output

In Table 14, the R^2 levels off in the 5-predictor range which is highlighted in ceramic yellow. This tells us that we do not need any more predictors (more than 5 predictors) in the model because unnecessary predictors are insignificant. Next, let's look at R_a^2, it increases and eventually it should go down. Between 3 and 4 predictor model range the R_a^2 decreases which is highlighted in red. This implies that 4 predictor model and the models after that are too big and thereby complexity of the model increases. Therefore, everything from 4 – predictor model is not considered as a potential best model. Let's check C_p now, it tells us that if there are any important predictors missing in the model however, it does not indicate if there are many predictors. Hence C_p is desired to be low however, if C(p) is approx. to p, also acceptable where p is parameters in the model. From Table 14, the C_p's highlighted in green are close to p or lower than p. Next is AIC, minimum AIC is in 2 – predictor model which is highlighted in purple. Finally, SBC, its minimum is in the 1 – predictor model range highlighted in fluorescent yellow.

Having used those criterions, it has helped us identify a few potentially best models. Here in Table 14, we look at the 1-predictor, 2-predictoe and 3-predictor models as they have the criteria metrics in the desired range and choose Models A, B, C, D and E. The models with no names are not considered for further analysis. From this analysis we know that A, B, C, D and E are potentially best models however, this analysis does not validate its significance level and VIF for high multicollinearity problem. This validation is done in 4.4.1 section of this report.

4.2 Stepwise Regression

In this method, the algorithm begins with no predictors. One-at-a-time, addition or delectation of predictor variable by identifying the smallest p – value and if p – value <= α_{in} or largest p – value and if p – value > α_{out} respectively until no predictors can be added or deleted. This model gives one potentially best model.

Now, let us obtain the one potential good model for the UTA students' model. The results of the Stepwise Regression are represented in Table 15 and 16.

Stepwise Selection: Step 2

Variable height Entered: R-Square = 0.4000 and C(p) = 1.8254

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6218.29614	3109.14807	29.33	<.0001
Error	88	9327.30841	105.99214		
Corrected Total	90	15546			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	28.23398	14.91725	379.69962	3.58	0.0617
gender	13.87007	2.51302	3228.77383	30.46	<.0001
height	0.18741	0.09254	434.73761	4.10	0.0459

Table 15: Stepwise Regression Final Step

All variables left in the model are significant at the 0.1000 level.

No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	gender		gender	1	0.3720	0.3720	3.8723	52.73	<.0001
2	height		height	2	0.0280	0.4000	1.8254	4.10	0.0459

Table 16: Summary of Stepwise Regression

For the UTA students' model, we can see that Stepwise Regression first brought in gender predictor variable. This matches with the best two - predictor model from the Best Subset Regression which is Model B. Since gender p-value < α = 0.1, it is not removed. Now out of the remaining predictors, the algorithm identified height to have the least p-value. Height p-value < α = 0.1 hence, height predictor variable was added into the model. Once the new model was obtained, both gender and height had p-value < α = 0.1 therefore, none of those predictor variables were removed. Next, out of the remaining predictor variable, none had p-vale < α = 0.1 and hence, the Stepwise Regression algorithm decided not to add anymore predictors into the model.

From this analysis, by satisfying the model search algorithm, we have one potential best model which is stated below in the equation 5, This model corresponds to the Model B in the Best Subset Regression.

$$Y_i = 28.23398 + 13.87007X_{i1} + 0.18741X_{i2} + \varepsilon_i$$

Equation 5: Model B

4.3 Backward Deletion

In this method we begin with the full model and regress Y on full set. Full set will have a set of p-values which is required for testing H_0: $\beta_k = 0$. Next, one-at-a-time we remove predictor variable with the largest p-value and rerun the regression which provides new p-values. The previous step is done until all the remaining predictors variables are significant at $\alpha = 0.1$. This method gives one potential good model.

Now, let us obtain the one potential good model for the UTA students' model. First, we begin with full model with all the predictors and interaction term i.e., gender, height, sleep time, water consumption, stdx2x3. Next regress Y on full model and we get Table 17 and 18.

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	28.23398	14.91725	379.69962	3.58	0.0617
gender	13.87007	2.51302	3228.77383	30.46	<.0001
height	0.18741	0.09254	434.73761	4.10	0.0459

Table 17: Last step of Backward Deletion

Summary of Backward Elimination								
Step	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	water_consumption	water_consumption	4	0.0013	0.4113	4.1946	0.19	0.6602
2	sleep_time	sleep_time	3	0.0016	0.4097	2.4262	0.23	0.6300
3	stdx2x3		2	0.0097	0.4000	1.8254	1.43	0.2358

Table 18: Summary of Backward Elimination

Now, we can see in Table 18 that water consumption, sleep time and stdx2x3 have been eliminated using the Backward Deletion algorithm. What this means is that gender and height together do a better job in predicting the weight of UTA students than water consumption, sleep time and stdx2x3. Therefore, we now have gender and height predictor variables.

From this analysis, by satisfying the model search algorithm, we have one potential best model which is stated below in the equation 6, This model corresponds to the Model B in the Best Subset Regression.

$$Y_i = 28.23398 + 13.87007X_{i1} + 0.18741X_{i2} + \varepsilon_i$$

Equation 6: Model B

4.4 Identifying potential best two model

4.4.1 Test for Significance and VISs

Now we have narrowed down to five potential best models i.e., A, B, C, D and E. Before we move forward with further analysis, we need to satisfy two conditions. First, all the predictors in the model should be significant at the $\alpha = 0.1$ level and second, the VIF should be less than 5 to guarantee multicollinearity is not a serious problem.

Each model i.e., A, B, C, D and E was regressed on Y and the following outputs were generated using SAS.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	58.24449	1.74552	33.37	<.0001	0
gender	gender	1	16.30368	2.24524	7.26	<.0001	1.00000

Table 19: Model A

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	28.23398	14.91725	1.89	0.0617	0
gender	gender	1	13.87007	2.51302	5.52	<.0001	1.29641
height	height	1	0.18741	0.09254	2.03	0.0459	1.29641

Table 20: Model B

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	58.15835	1.75422	33.15	<.0001	0
gender	gender	1	16.30763	2.25126	7.24	<.0001	1.00001
stdx2x3		1	0.85028	1.17325	0.72	0.4705	1.00001

Table 21: Model C

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	24.30322	15.24129	1.59	0.1144	0
gender	gender	1	13.56933	2.51960	5.39	<.0001	1.30951
height	height	1	0.21107	0.09442	2.24	0.0279	1.35617
stdx2x3		1	1.40103	1.17363	1.19	0.2358	1.04610

Table 22: Model D

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	31.68408	15.40629	2.06	0.0427	0
gender	gender	1	13.73665	2.51977	5.45	<.0001	1.30082
height	height	1	0.19718	0.09325	2.11	0.0373	1.31384
sleep_time	sleep_time	1	-0.74439	0.81850	-0.91	0.3656	1.01346

Table 23: Model E

From table 19 through 23 we can see that Model C, D and E each have one predictor which is not significant at $\alpha = 0.1$ (highlighted in red). Therefore, we can now rule out these three models for further analysis. Model A and B both have parameters that are significant at α = 0.1 and VIFs are just little more than one for each model (highlighted in green). Therefore, we can conclude that multicollinearity is not a serious problem.

4.4.2 Two Best Model

The two best models that satisfied both the conditions are:

$Y_i = 58.24449 + 16.303687X_{i1} + \varepsilon_i$ where X_{i1} is gender of UTA students

Equation 7: Model A

$Y_i = 28.23398 + 13.87007X_{i1} + 0.18741X_{i2} + \varepsilon_i$ where X_{i1} is gender and X_{i2} = height of UTA students

Equation 8: Model B

5. Model Selection

Now we have two potential best models that satisfied the significance and VIF conditions. Next, this section of the report analyses Model A and B further. For each model, model assumptions are verified, and the diagnostics are checked. This section is similar to the section 2.2 - Check for Model Assumption in the Preliminary Regression Analysis of this report.

5.1 Assumptions Verifications and Diagnostics Check on Model A

5.1.1 Check for constant variance and linear model assumption

Since Model A is a simple linear regression with only on parameter, we can check constant variance and linear model assumptions using e vs $\hat{Y}$ plot.

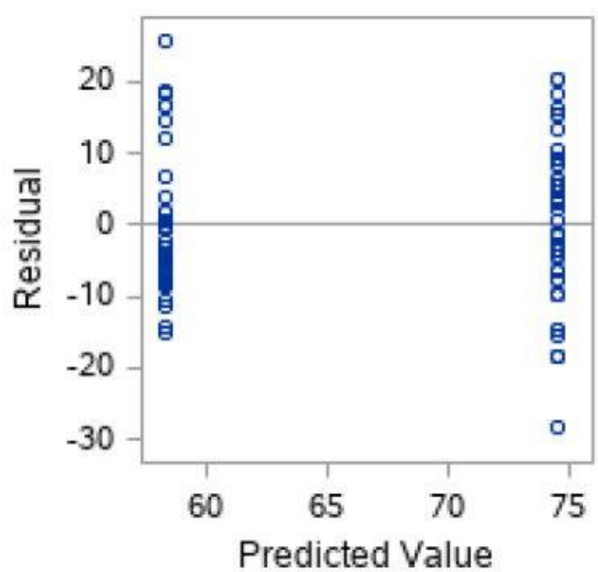

Figure 12: Model A residual plot

From Figure 12 we can say that there is no funnel shape therefore, constant variance assumption is satisfied for Model A. Regarding linear model assumption, there is no curvilinear shape in the residual plot which impels that the linear model assumption is also satisfied for Model A.

5.1.2 Check for normality assumption

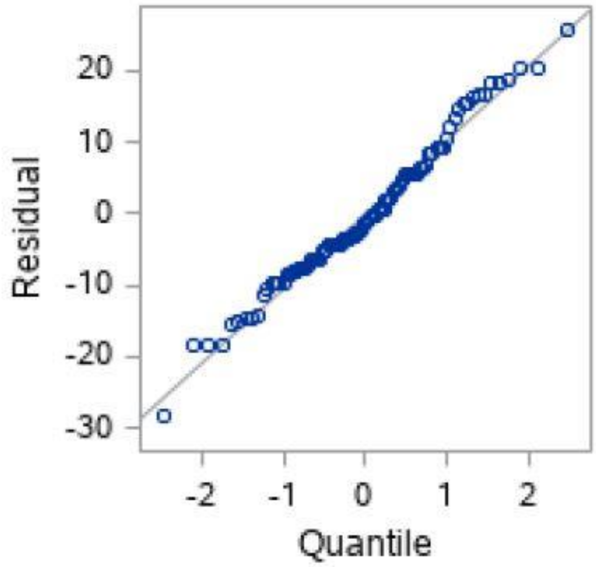

Figure 13: Model A NPP

From Figure 13 we can say that the NPP has a slight shorter right tail and shorter left tail than the normal distribution but, overall straight. Therefore, we can conclude that normality is close enough, although not perfect. Therefore, normality assumption is not satisfied for Model A.

5.1.3 Uncorrelation assumption

Regarding correlation assumption, the weight of UTA students is not collected in time order hence, time plots cannot be generated. Therefore, correlation assumption cannot be checked for Model A. Time series in not meaningful.

5.1.4 Model Diagnostics: Outliers, Leverage, and Influence

The necessary table to check the outliers is generated using SAS and we have cutoff values from Table 24, tinvtres is compared with R Student and finv50 is compared with CooksDi to identify outliers and its influence on the model respectively . Cuttoff for h_{ii} , DFFITS and DFBETAS is found to be $\frac{2p}{n} = \frac{2*2}{91} = 0.04395$, $2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{2}{91}} = 0.2964$ and $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{91}} = 0.2096$ respectively.

Obs	tinvtres	finv50
1	3.37528	0.69857

Table 24: Cutoff values

Every observation in the Cooks Distance does not exceed 0.69857. Cooks distance does not detect influence. Similarly, every observation in R Student does not exceed 3.37931. Hence, we can conclude that there are no y-outliers. Also, every observation in the h_{ii} does not exceed 0.04395 therefore, we can conclude that there are no outliers in the x-direction. Since there are no outliers, the corresponding |DFFITS| and |DFBETAS| cannot be checked for possible influence on the fitted value ($\hat{y}_i$) and LSCs. In the interest of page limit for this report, since no outliers were identified, the output Table 34 from SAS for outliers are attached in the appendix section 9.2 of this report.

5.1.5 Variance Inflation Factor

Model A has only one predictor, the VIF will be equal to 1. Since there no more than one predictor variable, correlation among predictors is not a problem therefore, check for VIF in Model A is not meaningful.

5.2 Assumptions Verifications and Diagnostics Check on Model B

5.2.1 Check for constant variance

For Model B we can check only constant variance using e vs $\hat{Y}$ plot.

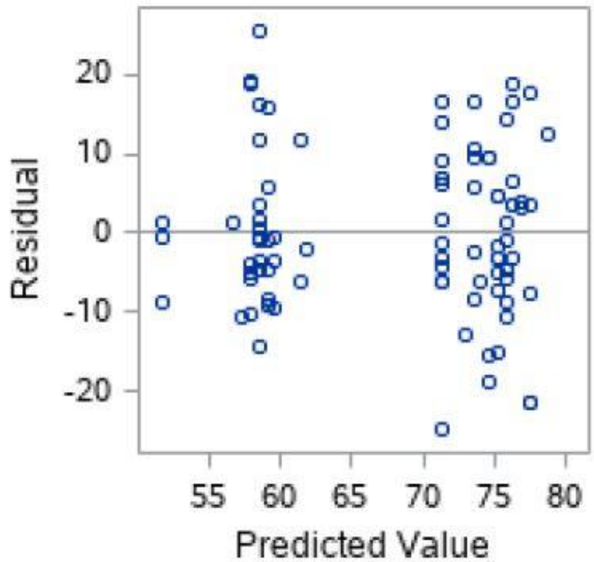

Figure 14: Model B residual plot

From Figure 14 we can say that there is no funnel shape therefore, constant variance model assumption is satisfied for Model B.

5.2.2 Check for model form

To check the model form of Model B we need to review two plots residual vs gender and residual and height plots simultaneously. From Figure 15 we can conclude that Model B form is adequate because there is no curvature in either of the two plots.

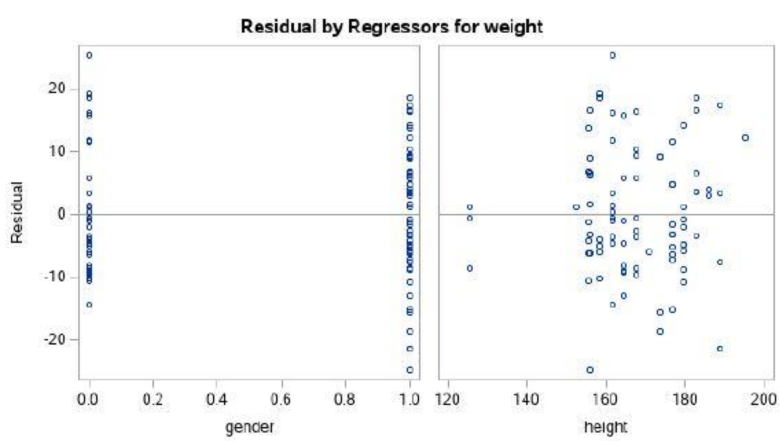

Figure 1511: e vs gender and e vs height plots

5.2.3 Check for normality assumption

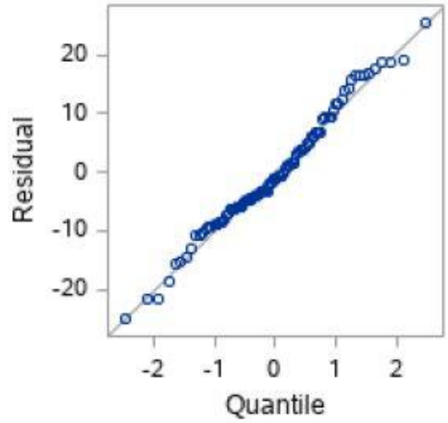

Figure 126: Model B NPP

From Figure 16 we can say that the NPP has a very slight shorter right tail than the normal distribution but, overall, it's pretty straight. Therefore, we can conclude that normality is ok for Model B. Normality is satisfied.

5.2.4 Uncorrelation assumption

Regarding correlation assumption, the weight of UTA students is not collected in time order hence, time plots cannot be generated. Therefore, correlation assumption cannot be checked for Model B. Time series in not meaningful.

5.2.5 Model Diagnostics: Outliers, Leverage, and Influence

The necessary cutoff values are as follows for Table 26 below which is generated using SAS and cutoff values from Table 25, tinvtres is compared with R Student and finv50 is compared with CooksDi to identify outliers and its influence on the model respectively . Cuttoff for h_{ii} , DFFITS and DFBETAS is found to be $\frac{2p}{n} = \frac{2*3}{91} = 0.0659$, $2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{3}{91}} = 0.3631$ and $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{91}} = 0.2096$ respectively.

Obs	tinvtres	finv50
1	3.37660	0.79481

Table 25: Cutoff values

Output Statistics									
Obs	Residual	Cooks Di	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS		
							Intercept	Gender	Height
5	-8.7118	3.93E-02	-0.9042	0.1259	1.1512	-0.3232	-0.3195	-0.0348	0.3030
22	1.2882	0.00086	1.33E-01	0.1259	1.1833	0.0505	0.0470	0.0051	-0.0446
35	-0.6827	2.42E-04	-0.0705	0.1259	1.1838	-0.0268	-0.0249	-0.0027	0.0238

Table 26

In the best interest of space, the output is provided for the observations with outliers only. Every observation in the Cooks Distance does not exceed 0.79481. Cooks distance does not detect influence. Similarly, every observation in R Student does not exceed 3.37660. Hence, we can conclude that there are no y-outliers. Observations 5, 22 and 35 are greater than 0.0659 therefore, we can conclude that these observations are outliers in the x-direction. Next, the corresponding |DFFITS| for x-outliers is compared with cutoff 0.3631 and no observation is greater than the cutoff value therefore, we can conclude that these observations does not influence the fitted value ($\hat{y}_i$). Finally, the corresponding |DFBETAS| is compared with $\frac{2}{\sqrt{n}} = 0.2096$ and observations 5 than 0.2096 therefore, we can conclude that these observations influence the LSEs.

5.2.6 Variance Inflation Factor

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	1	28.23398	14.91725	1.89	0.0617	0
gender	gender	1	13.87007	2.51302	5.52	<.0001	1.29641
height	height	1	0.18741	0.09254	2.03	0.0459	1.29641

Table 27: Parameter Estimates

Moving on with the model diagnostics, next, we need to assess how sever multicollinearity is. To assess this, VIF is calculated using SAS and the output is represented in Table 27. From Table 27 we firstly have, $(VIF)_1$ = 1.29641 where gender is regressed on height. Secondly, we have, $(VIF)_2$ = 1.29641 where height is regressed on gender. Next, we need to find $(VIF)_k$ value by using the formula $(\bar{V}\bar{I}\bar{F}) = \frac{\sum_{k=1}^{p-1}(VIF)_1}{p-1} = \frac{1.29641+1.29641}{2-1} = 2.59282$

Since $(VIF)_k$ = 2.59282 and max (VIF) = 1.29641 which is little more than 1.0, we can conclude that serious multicollinearity is not a problem for Model B.

5.3 Final Model Selection

In this section of the report, the final model is selected. Model A and Model B both are fully discusses with regards to assumptions, diagnostics, etc., that were discussed earlier in this report by creating a table with the list of pros and cons of each model, we can justify and select one best model.

Final Model Selection				
Parameters	**Model A:** $Y_i = 58.24449 + 16.303687X_{i1} + \varepsilon_i$		**Model B:** $Y_i = 28.23398 + 13.87007X_{i1} + 0.18741X_{i2} + \varepsilon_i$	
	Pros	**Cons**	**Pros**	**Cons**
C(p)		C(p) is little higher than p	Has the least C(p) value and is lower than p	
AIC			Has the minimum AIC vale	
SBC	Has the minimum SBC value		Has SBC very close to the minimum value	

R^2		Has comparatively lower R^2 at 37.20%	Has better R^2 at 40%	
Outliers	No outliers			Has 3 outliers
Influence	No influential outliers			Outliers are influential on individual LSCs
Multicollinearity	Is not a problem		Is not a problem	
Normality		Normality not satisfied	Normality is satisfied	
Other model assumptions	Satisfied		Satisfied	
SSR		Comparatively, SSR is lower	SSR is higher	
SSE		Comparatively, SSE is higher	SSE is lower	
Complexity	Less complex with only one predictor			Comparatively more complex with two predictors

Table 28

Clearly, we can see from Table 28 that Model B outperforms Model A. In addition to different criterions and its metrics, we can say that height and gender together play a major role in determining the weight of a student. This can be explained by asking a question, and the statical way of asking it is, does gender and height factors together contribute to variability in the weight response of UTA students? Answer to this question is according to the National Institute of Diabetes and Digestive and Kidney Diseases a part of the U.S. Department of Health and Human Services, obesity is more common in woman than men because of where (location) the body stores fat [1] and as a person's height increases, their body volume also increases which results in increase in mass i.e., weight. From this explanation we can say that Model A provides only the mean weight for male and female UTA students however, Model B provides a more accurate weight for each individual UTA student by taking gender and height into consideration. Hence, Model B is more accurate than Model A i.e., gender and weight together do a better job in explaining the weight of a UTA student. Based on pros and cons, along with the explanation we can say that Model B is the best overall model.

6. Final Multi Linear Regression Model

Now that we have our final best overall model, this section of the report presents and interprets its meaning. Model B is represented in its equation form in equation 9 below where x_{i1} is gender, x_{i2} is height and Y_i is the weight of UTA students.

$$Y_i = 28.23398 + 13.87007X_{i1} + 0.18741X_{i2} + \varepsilon_i$$

Equation 9: Overall best model

6.1 Interpretation of the best model

6.1.1 Parameter Estimates

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	28.23398	14.91725	1.89	0.0617	0
gender	gender	1	13.87007	2.51302	5.52	<.0001	1.29641
height	height	1	0.18741	0.09254	2.03	0.0459	1.29641

Table 29: Model B Parameter Estimates

From Table 29, the y-intercept and slope for x1 and x2 are found to be 28.23398, 13.87007 and 0.18741 respectively. With this, the equation of the model can be stated as $\mathbf{\hat{y}_{i = 1 \text{ to } 91} = 28.23398 + 13.87007\ x_{i1 = 1 \text{ to } 91} + 0.18741\ x_{i2 = 1 \text{ to } 91}}$, here $\hat{y}$ represents the estimated mean weight of a student, X_{i1} and x_{i2} represent gender and weight of UTA students respectively.

The slopes in the equation of the line states that if a student gender is male ($x_{i1} = 1$) then his weight will increase by 13.87007 kg and if a student is female ($x_{i1} = 0$), then, her weight will not increase by 13.87007 kg. Next, for every one-centimetre increase in height of a UTA student their corresponding weight will increase by 0.18741 kg. The y-intercept is irrelevant because if both $x_{i1} = 0$ and $x_{i2} = 0$ i.e., height is 0 cm the weight will be 28.23398 kg which is impossible and does not make any sense.

6.1.2 ANOVA Table

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6218.29614	3109.14807	29.33	<.0001
Error	88	9327.30841	105.99214		
Corrected Total	90	15546			

Table 30: Model B ANOVA Table

Table 30 represents the Analysis of Variance for Model B.

From Table 30 we have the sum of square regression (SSR) equal to 6218.29614 which is the deviation of predicted weights ($\hat{y}$) from the mean weight ($\bar{y}$) of all 91 students explained by the model. The sum of square error (SSE) is equal to 9327.30841 and it is the deviation of the observed weight (y) from the mean weight ($\bar{y}$) of all 91 students which is not explained by the model and considered as errors. The sum of square total (SST) is the sum of the deviation of the predicted weight ($\hat{y}$) and the sum of the deviation of observed weight (y) from the mean weight ($\bar{y}$) of all 91 students which is found out to be 15546.

Now we can check what percentage of total variation from mean weight of UTA students is explained by this Model B by calculating coefficient of determination, R^2 is found out to be 0.39999999 approximately 40%. This implies that only 40% of variation in weight of the UTA students is explained by the student's gender and height combined.

With the ANOVA table we can also test the significance of regression using F-Test.

H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots\ldots\ \beta_{p-1} = 0$ which means reduced model, $Y_i = \beta_0 + \varepsilon_i$
H_1: at least one β_k is not zer0 which means full model, $Y_i = x\beta + \varepsilon_i$
CI = 90% and $\alpha = 0.1$
From ANOVA table we have $F^* = 29.33$
F (1-α; p-1; n-p) = F (0.9; 2; 88) = 3.1[3]
Since F^* > F (1-α; p-1; n-p i.e., 29.33 > 3.1 we reject H_0. This is a strong conclusion. We can conclude that regression is significant.

6.1.3 Inferences

We are particularly interested in studying the case for male gender ($x_1 = 1$) and height 183 cm ($x_2 = 183$) because, all the members of the group for this project are male students of UTA and their average height is 183 cm. Therefore, x_h^T [1 1 183]. By regressing weight on gender and height we have, Table 31 and 32.

b
28.233981
13.870072
0.1874133

Table: 31

X'X Inverse, Parameter Estimates, and SSE					
Variables	**Label**	**Intercept**	**Gender**	**Height**	**Weight**
Intercept	Intercept	2.099443004	0.14021747	-0.012937391	28.23398131
Gender	Gender	0.14021747	0.059582648	-0.001049177	13.87007191
Height	Height	-0.012937391	-0.001049117	0.000080793	0.187413344
Weight	Weight	28.23398131	13.87007191	0.187413344	9327.308412

Table: 32

6.1.3.1 Confidence Interval for Slopes

Using Table 30, 31 and 32 we can find $S\{b_k\} = \sqrt{MSE\ k^{th} diagonal\ element\ of\ (x^T x)^{-1}]}$ we have $S\{b_1\}$ = 2.5130 and $S\{b_2\}$ = 0.09253

Using 90% CI we have $\alpha = 0.1$, now t (1 - $\frac{\alpha}{2}$; n-p) = t (0.95;88) = 1.6623[4].

CI for Slopes is given by the formula $b_k \pm$ t (1 - $\frac{\alpha}{2}$; n-p) * $S\{b_k\}$

CI for gender slope is found out to be (9.6927; 18.047). We are 90% confident that the change in mean weight of a UTA student increases between 9.6927 kg and 18.047 kg if the student is male.

CI for height slope is found out to be (0.0336; 0.34122). With 90% confidence we can say that change in mean weight of a UTA student increases between 0.0336 kg and 0.34122 kg for every unit cm in increase of height.

6.1.3.2 Confidence Interval for y – intercept

From the survey we have the minimum height of a UTA student, which is 125 cm therefore, X = 0 is outside of the X- range. Hence y – intercept is not meaningful and determining a confidence interval for the y - intercept is irrelevant because if both $x_{i1} = 0$ and $x_{i2} = 0$ i.e., height is 0 cm the weight will be 28.23398 kg which is impossible and does not make any sense.

6.1.3.3 Confidence Interval for Mean Response

Using Table 30, 31 and 32 we can find $\hat{Y}_h = X_h^T * b$ we have $\hat{Y}_h$ = 76.4006kg. Next, using $S\{\hat{Y}h\} = \sqrt{MSE\ [x_h^T((x^Tx)^{-1}x^h]}$ we have $S\{\hat{Y}h\}$ = 1.35591.

Using 90% CI we have α = 0.1, now $t(1 - \frac{\alpha}{2}; n\text{-}p) = t(0.95;88)$ = 1.6623[4].

CI for mean response is given by the formula $\hat{Y}_h \pm t(1 - \frac{\alpha}{2}; n\text{-}p) * S\{\hat{Y}h\}$ = (74.1466; 78.6545).

From the calculated CI we can say that we are 90% confidence that mean weight of UTA students with gender male and height 183 cm will lie between 74.1466kg and 78.6545kg.

6.1.3.4 Prediction Interval for New Response

Using Table 30, 31 and 32 we have $\hat{Y}_h$ = 76.4006kg. Next, using $S\{pred\} = \sqrt{(S\{\hat{Y}h\})^2 + MSE}$ we have S{pred} = 10.3841

Using 90% CI we have α = 0.1, now $t(1 - \frac{\alpha}{2}; n\text{-}p) = t(0.95;88)$ = 1.6623[4].

CI for mean response is given by the formula $\hat{Y}_h \pm t(1 - \frac{\alpha}{2}; n\text{-}p) * S\{pred\}$ = (59.1319; 93.6620).

From the calculated CI we can say that we are 90% confidence that the actual weight of the next male UTA student with height of 183 cm will lie between 59.1319 kg and 93.6620 kg.

6.1.3.4 Joint CI for parameters

Now we have CI = 90%, α = 0.1, p = 3, n = 91, g = 2, $S\{b_1\}$ = 2.5130 and $S\{b_2\}$ = 0.09253

Two-sided joint CI is given by $b_k \pm B\ S\{b_k\}$ where $B = t(1 - \frac{\alpha}{2g}; n\text{-}p) = t(0.975; 88)$ = 1.9873[4]

Using two-sided join CI β_1 was found out to be (8.8759, 18.86415) and β_2 was found out to be (0.003528, 0.371298). From this we can say that we are 90% confidence that β_1 is in (8.8759, 18.86415) and β_2 is in (0.003528, 0.371298) simultaneously.

6.1.3.5 Confidence Region for the entire regression surface

We above calculations CI = 90%, α = 0.1, p = 3, n = 91, n - p = 88 and $\hat{Y}_h$ = 76.4006 kg

CR with limits is $\hat{Y}_h \pm W\ S\{\hat{Y}h\}$ where $W = p\ F(1\text{-}\alpha; p\text{-}n) = 3 * F(0.9;3;88)$ = 2.5379[3]

Therefore, CR for the entire regression surface is (72.959, 79.841) 90% CI

7. Final discussion

7.1 Summary

In this project, the preliminary model was first introduced. Its assumptions were checked, and all the assumptions were satisfied. Necessary test and diagnostics were conducted such as, Modified Levene test, test for normality and Bonferroni outlier test to check these assumptions. After this, the model had 4 predictors i.e., gender, height, sleep time and water consumption. Next, interaction terms were explored. By studying the Partial Regression Plots it was found that x2x3 interaction term can be added however, correlation among interaction terms and original predicts were very high. Hence, the standardized form of interaction term was added into the model. After adding interaction term, the model had 5 predictors, gender, height, sleep time, water consumption and stdx2x3. Next, using Best Subset Regression, Backward Deletion and Stepwise Regression model search method, we were able to come up with 5 potential best models, Model A, B, C, D, and E. Out of these 5 models, only Model A and B satisfied the significance and VIF conditions. Model A was one-predictor model and Model B was a two-predictor model. To select the best model among Model A and B, the assumptions and diagnostics were verified. Pros and cons table was created to help in identifying the overall best model and Model B was selected as the overall best model. Next, we presented and interpreted the final model. We conducted the F-Test for to check for significance and finally, we discussed the fit of the model and interpreted inferences.

7.2 Future work

Since only 40% of the variability is explained by the final model (Model B) there is room for improvement. During the survey, the data was self-reported by the students and for the analysis of this model, it was assumed that the self-reported data is accurate. To increase the accuracy of the model, during survey, measuring devices such as measuring tape and weighing scale is recommended to be used to take accurate readings. Also, additional predictor variables such as calories consumed, sugar level in the body, body metabolism rate and blood pressure may be considered.

8. References

[1] “Factors affecting weight and health”, *National Institute of Diabetes and Digestive and Kidney Diseases,* Factors Affecting Weight & Health | NIDDK (nih.gov).

[2] McCall Minnor “7 Things that affect your weight (besides diet and exercise)”, *Aaptiv,* 7 Things That Affect Your Weight (Besides Diet and Exercise) - Aaptiv.

[3] F – Distribution: 1.3.6.7.3. Upper Critical Values of the F Distribution (nist.gov)

[4] T – Distribution: T Value (Critical Value) Calculator (meracalculator.com)

9. Appendix

9.1 About Survey

The data for SLR & MLR project was collected using Microsoft forms. The title of survey is "Factors affecting the weight of humans."

This is a completely anonymous survey that is being conducted for a group project. The sole purpose of the information used from the data collected in this form is to analyze and study the factors that affect human weight. The data collected from this survey is completely anonymous and will not be published.

9.1.1 Data collection

Firstly, a survey form was drafted online and the barcode to navigate to the same was generated. Students in the vicinity of MAC (maverick activities center) and the University center (UC) were approached and requested to help with the data collection process. The data collected was not measured, it was self-reported by the students themselves. For the purpose of this analysis, it is assumed that the self-reported data provided by the students is accurate.

9.1.2 Question

As part of the survey, the following questions were asked to the participants to collect the data.

1. Do you consent to take this survey?
2. What is your gender?
3. How tall are you? (Feet and inches)
4. What is your weight? (Kg or lbs)
5. On an average how many hours do you sleep in a day?
6. Approximately how much water do you consume in a day? (Litres or oz or gallons)

The responses received from 91 UTA students to these questions are tabulated and presented as raw data in section 9.1.3 of this report.

9.1.3 Raw Data

Table 33: Raw Survey Data

N	Gender	Height (cm)	Weight (Kg)	Sleep (Hr)	Water Consumption (Lt)
1	Male	156	65.00	7.5	5.00
2	Male	183	80.00	7	2.00
3	Female	168	56.00	7	3.00
4	Female	158	52.00	7.5	2.00
5	Female	125	43.00	6	1.00
6	Male	155	67.00	6	2.00
7	Female	158	53.00	7	1.00
8	Male	189	56.00	6	3.00
9	Male	189	56.00	6	3.00
10	Male	168	84.00	6	3.00
11	Male	156	46.40	9	2.00
12	Male	177	68.00	6.5	8.00
13	Female	158	76.66	6	4.00
14	Female	162	62.00	6	4.00
15	Male	177	72.00	5	3.00
16	Male	177	80.00	7	3.00

17	Male	156	78.00	1.5	3.00
18	Male	177	73.56	8	2.50
19	Female	158	77.11	3	1.00
20	Female	162	84.00	8	2.00
21	Female	162	57.61	7	2.66
22	Female	125	53.00	5.5	3.00
23	Male	174	84.00	6	5.00
24	Male	189	81.00	9	4.00
25	Male	168	71.00	6.5	3.00
26	Male	156	88.00	6.5	1.89
27	Male	186	81.00	6.5	3.00
28	Male	183	73.00	7	2.00
29	Male	180	90.00	7	2.00
30	Male	177	80.00	6	2.00
31	Male	155	85.00	7	2.00
32	Male	180	67.00	8	2.50
33	Female	162	54.00	8	3.00
34	Female	162	59.00	6	1.00
35	Female	125	51.03	7	2.00
36	Male	180	71.00	7	3.00
37	Male	174	56.00	6.5	3.00
38	Male	156	77.50	5	3.00
39	Female	162	55.00	7.5	8.00
40	Female	152	58.00	8	1.00
41	Female	165	54.50	7	1.00
42	Male	180	77.00	9	1.00
43	Male	156	80.29	8	4.00
44	Female	162	60.00	7	5.00
45	Female	168	58.97	6.5	7.57
46	Male	183	92.99	6	3.00
47	Male	156	68.04	5.5	15.14
48	Male	168	79.38	7.5	11.36
49	Female	158	47.63	6.5	0.59
50	Male	177	70.00	6	5.00
51	Male	180	75.00	7	4.00
52	Male	156	73.00	8	3.00
53	Male	195	91.00	6.5	2.00
54	Male	189	95.00	8	3.00
55	Female	165	50.80	7	1.00
56	Female	162	74.84	6	3.79
57	Female	165	74.84	7	2.07
58	Male	177	60.00	8	2.00
59	Female	177	55.00	5	2.00
60	Male	183	95.00	6.5	6.00
61	Female	162	70.31	8	1.89
62	Female	165	49.90	7	1.48
63	Male	189	70.00	6	4.00
64	Female	155	46.72	6.5	2.10

65	Female	165	49.90	8	2.50
66	Female	165	50.00	7	2.00
67	Male	180	71.00	7	2.00
68	Male	174	59.00	7	4.00
69	Male	168	83.00	7.5	6.00
70	Female	168	50.00	7	1.00
71	Female	165	65.00	7	3.00
72	Male	155	65.00	1	8.00
73	Female	177	73.00	5.5	7.00
74	Female	165	58.00	8	3.00
75	Male	183	80.00	7	7.00
76	Male	177	72.00	7	2.00
77	Male	180	70.00	7	3.00
78	Male	180	65.00	5	2.00
79	Male	171	68.00	9	3.00
80	Female	162	44.00	6	3.00
81	Female	162	58.00	6.5	3.00
82	Male	165	60.00	8.5	8.00
83	Male	168	90.00	6	2.00
84	Male	174	84.00	8	2.00
85	Male	155	70.00	9	2.00
86	Male	183	83.00	6	5.00
87	Male	168	65.00	6.5	1.50
88	Male	155	78.00	7.5	9.00
89	Female	158	54.00	6	1.00
90	Female	180	60.00	8.5	4.00
91	Male	186	80.00	5.5	2.00

9.2 Model A Outliers Table SAS Output

Table 34: Model A Outliers Table SAS Output

Output Statistics								
Obs	**Residual**	**CooksDi**	**RStudent**	**Hat Dlag H**	**Cov Ratio**	**DFFITS**	**DFBETAS**	
							Intercept	**Gender**
1	-9.5482	0.007839	-0.9193	0.0182	1.0221	-0.125	0.0000	-0.0787
2	5.4518	0.002556	0.5232	0.0182	1.0353	0.0712	0.0000	0.0448
3	-2.2445	0.000675	-0.2162	0.0278	1.051	-0.037	-0.0365	0.0284
4	-6.2445	0.005224	-0.6025	0.0278	1.0435	-0.102	-0.1018	0.0792
5	-15.2445	0.031132	-1.4862	0.0278	1.0012	-0.251	-0.2512	0.1953
6	-7.5482	0.004899	-0.7254	0.0182	1.0294	-0.099	0.0000	-0.0621
7	-5.2445	0.003685	-0.5057	0.0278	1.046	-0.086	0.0000	0.0665
8	-18.5482	0.029580	-1.8101	0.0182	9.684	-0.246	0.0000	-0.1549
9	-18.5482	0.029580	-1.8101	0.0182	9.684	-0.246	0 0000	-0.1549
10	9.4518	0.007681	0.9099	0.0182	1.0225	0.1238	0.0000	0.0779

11	-28 1482	0.068123	-2.8161	0.0182	0.8767	-0.383	0.0000	-0.2410
12	-6.5482	0.003687	-0.6289	0.0182	1.0325	-0.086	0.0000	-0.0538
13	18.4126	0.045417	1.8055	0.0278	0.9783	0.3052	0.3052	-0.2373
14	3.7555	0.001889	0.3619	0.0278	1.049	0.0612	0.0612	-0.0476
15	-2.5482	0.000558	-0.2442	0.0182	1.0404	-0.033	0.0000	-0.0209
16	5.4518	0.002556	0.5232	0.0182	1.0353	0.0712	0.0000	0.0448
17	3.4518	0.001024	0.331	0.0182	1.0392	0.045	0.0000	0.0283
18	-0.9882	0.000084	-0.0947	0.0182	1.0416	-0.013	0.0000	-0.0081
19	18.8661	0.047682	1.8517	0.0278	0.9746	0.313	0.3130	-0.2433
20	25.7555	0.088864	2.5715	0.0278	0.9102	0.4347	0.4347	-0.3379
21	-0.6383	0.000055	-0.0615	0.0278	1.052	-0.01	-0.0104	0.0081
22	-5.2445	0.003685	-0.5057	0.0278	1.046	-0.086	-0.0855	0.0665
23	9.4518	0.007681	0.9099	0.0182	1.0225	0.1238	0.0000	0.0779
24	6.4518	0.003579	0.6196	0.0182	1.0328	0.0843	0.0000	0.0530
25	-3.5482	0.001082	-0.3402	0.0182	1.0391	-0.046	0.0000	-0.0291
26	13.4518	0.015558	1.3013	0.0182	1.0028	0.1771	0.0000	0.1114
27	6.4518	0.003579	0.6196	0.0182	1.0328	0.0843	0.0000	0.0530
28	-1.5482	0.000206	-0.1484	0.0182	1.0413	-0.02	0.0000	-0.0127
29	15.4518	0.020528	1.4994	0.0182	0.9905	0.204	0.0000	0.1283
30	5.4518	0.002556	0.5232	0.0182	1.0353	0.0712	0.0000	0.0448
31	10.4518	0.009392	1.0072	0.0182	1.0182	0.1371	0.0000	0.0862
32	-7.5482	0.004899	-0.7254	0.0182	1.0294	-0.099	0.0000	-0.0621
33	-4.2445	0.002413	-0.4091	0.0278	1.0481	-0.069	-0.0692	0.0538
34	0.7555	0.000076	0.0728	0.0278	1.052	0.0123	0.0123	-0.0096
35	-7.2154	0.006974	-0.6967	0.0278	1.0406	-0.118	-0.1178	0.0916
36	-3.5482	0.001082	-0.3402	0.0182	1.0391	-0.046	0.0000	-0.0291
37	-18.5482	0.029580	-1.8101	0.0182	0.9684	-0.246	0.0000	-0.1549
38	2.9518	0.000749	0.283	0.0182	1.0399	0.0385	0.0000	0.0242
39	-3.2445	0.001410	-0.3126	0.0278	1.0497	-0.053	-0.0528	0.0411
40	-0.2445	0.000008	-0.0235	0.0278	1.0521	-0.004	-0.0040	0.0031
41	-3.7445	0.001878	-0.3608	0.0278	1.049	-0.061	-0.0610	0.0474
42	2.4518	0.000517	0.235	0.0182	1.0405	0.032	0.0000	0.0201
43	5.7376	0.002830	0.5507	0.0182	1.0347	0.0749	0.0000	0.0471
44	1.7555	0.000413	0.1691	0.0278	1.0514	0.0286	0.0286	-0.0222
45	0.7225	0.000070	0.0696	0.0278	1.052	0.0118	0.0118	-0.0091
46	18.4382	0.029230	1.7989	0.0182	0.9692	0.2448	0.0000	0.1540
47	-6.5094	0.003643	-0.6251	0.0182	1.0326	-0.085	0.0000	-0.0535
48	4.8304	0.002006	0.4634	0.0182	1.0367	0.0631	0.0000	0.0397
49	-10.6173	0.015101	-1.0285	0.0278	1.0272	-0.174	-0.1738	0.1352
50	-4.5482	0.001779	-0.4363	0.0182	1.0373	-0.059	0.0000	-0.0373
51	0.4518	0.000018	0.0433	0.0182	1.0418	0.0059	0.0000	0.0037
52	-1.5482	0.000206	-0.1484	0.0182	1.0413	-0.02	0.0000	-0.0127
53	16.4518	0.023271	1.5992	0.0182	0.9838	0.2176	0.0000	0.1369
54	20.4518	0.035963	2.0039	0.0182	0.9529	0.2727	0.0000	0.1715
55	-7.4422	0.007420	-0.7187	0.0278	1.0398	-0.122	-0.1215	0.0944

56	16.5982	0.036907	1.622	0.0278	0.9919	0.2742	0.2742	-0.2131
57	16.5982	0.036907	1.622	0.0278	0.9919	0.2742	0.2742	-0.2131
58	-14.5482	0.018197	-1.4097	0.0182	0.9963	-0.192	0.0000	-0.1207
59	-3.2445	0.001410	-0.3126	0.0278	1.0497	-0.053	-0.0528	0.0411
60	20.4518	0.035963	2.0039	0.0182	0.9529	0.2727	0.0000	0.1715
61	12.0623	0.019491	1.1705	0.0278	1.0201	0.1979	0.1979	-0.1538
62	-8.3494	0.009339	-0.8069	0.0278	1.0367	-0.136	-0.1364	0.1060
63	-4.5482	0.017790	-0.4363	0.0182	1.0373	-0.059	0.0000	-0.0373
64	-11.5245	0.017792	-1.1176	0.0278	1.0228	-0.189	-0.1889	0.1469
65	-8.3494	0.009339	-0.8069	0.0278	1.0367	-0.136	-0.1364	0.1060
66	-8.2445	0.001410	-0.7967	0.0278	1.0371	-0.135	-0.1347	0.1047
67	-3.5482	0.035963	-0.3402	0.0182	1.0391	-0.046	0.0000	-0.0291
68	-15.5482	0.019491	-1.509	0.0182	0.9899	-0.205	0.0000	-0.1292
69	8.4518	0.009339	0.8129	0.0182	1.0263	0.1106	0.0000	0.0696
70	-8.2445	0.017790	-0.7967	0.0278	1.0371	-0.135	-0.1347	0.1047
71	6.7555	0.017792	0.6521	0.0278	1.042	0.1102	0.1102	-0.0857
72	-9.5482	0.009339	-0.9193	0.0182	1.0221	-0.125	0.0000	-0.0787
73	14.7555	0.009106	1.4374	0.0278	1.0044	0.243	0.2430	-0.1889
74	-0.2445	0.001082	-0.0235	0.0278	1.0521	-0.004	-0.0040	0.0031
75	5.4518	0.020785	0.5232	0.0182	1.0353	0.0712	0.0000	0.0448
76	-2.5482	0.006142	-0.2442	0.0182	1.0404	-0.033	0.0000	-0.0209
77	-4.5482	0.009106	-0.4363	0.0182	1.0373	-0.059	0.0000	-0.0373
78	-9.5482	0.006114	-0.9193	0.0182	1.0221	-0.125	0.0000	-0.0787
79	-6.5482	0.007839	-0.6289	0.0182	1.0325	-0.086	0.0000	-0.0538
80	-14.2445	0.029167	-1.3865	0.0278	1.0076	-0.234	-0.2344	0.1822
81	-0.2445	0.000008	-0.0235	0.0278	1.0521	-0.004	-0.0040	0.0031
82	-14.5482	0.002556	-1.4097	0.0182	0	-0.192	0.0000	-0.1207
83	15.4518	0.000558	1.4994	0.0182	0.9905	0.204	0.0000	0.1283
84	9.4518	0.001779	0.9099	0.0182	1.0225	0.1238	0.0000	0.0779
85	-4.5482	0.007839	-0.4363	0.0182	1.0373	-0.059	0.0000	-0.0373
86	8.4518	0.003687	0.8129	0.0182	1.0263	0.1106	0.0000	0.0696
87	-9.5482	0.027182	-0.9193	0.0182	1.0221	-0.125	0.0000	-0.0787
88	3.4518	0.000008	0.331	0.0182	1.0392	0.045	0.0000	0.0283
89	-4.2445	0.018197	-0.4091	0.0278	1.0481	-0.069	-0.0692	0.0538
90	1.7555	0.020528	0.1691	0.0278	1.0514	0.0286	0.0286	-0.0222
91	5.4518	0.007681	0.5232	0.0182	1.0353	0.0712	0.0000	0.0448