Problem 1

(1) Performing exploratory analysis on the above variables.

See the following charts:

"Box plot for FEV1_mother by Height_mother",

"Scatter plot for FEV1_mother vs. Height_mother",

"Normal distribute plot for Height_mother",

"Normal distribute plot for FEV1_mother".

(2) Performing regression analysis.

See the following charts:

"Simple regression for FEV1_mother vs. Height_mother",

"Univariate analysis for FEV1_mother vs. Height_mother".

(3) Would you apply any transformation to the "height" variable? What kind of transformation if any?

The log transformation. We can transform the height into log(height).

See "Simple regression for FEV1_mother vs. logHeight_mother".

As we see, the result of these two regression are similar, including the residual, F-value and R-square. As a result, we do not need to apply the log transformation to "height" variable.

(4) Is the regression line a good predictor? Why or why not?

Yes. This regression line is a good predictor.

The F value is 40.73 and Pr > F is <0.0001, which is highly significant.

The R-Square is 0.2158, which indicates how well the data fit this model.

The Pr > |t| for intercept and Height_mother are 0.002 and <0.0001, which are also significant.

The model is FEV1_mother = Height_mother × 9.17 – 290.34.

Simple Regression for FEV1_mother vs. Height_mother

The REG Procedure
Model: MODEL1
Dependent Variable: FEV1_mother

Number of Observations Read	150
Number of Observations Used	150

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	76391	76391	40.73	<.0001
Error	148	277592	1875.61829		
Corrected Total	149	353982			

Root MSE	43.30841	R-Square	0.2158
Dependent Mean	297.31333	Adj R-Sq	0.2105
Coeff Var	14.56659		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-290.34463	92.15023	-3.15	0.0020
Height_mother	1	9.16878	1.43669	6.38	<.0001

(5) What are high leverage and influential observations? Why?

leverage:

The highest five leverage points are shown in the following chart, which are 88, 19, 4, 15 and 3. These points' height and FEV1 are: (88): 57, 221; (19): 57, 204; (4): 58, 206; (15): 59, 257; (3) 59, 265.

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.00667625	147	0.0352154	3
0.00667625	144	0.0352154	15
0.00667625	142	0.0475261	4
0.00667625	141	0.0620378	19
0.00667625	140	0.0620378	88

influential observation:

cookd and dffits:

The highest five cookd and dffits points are shown in the following charts. Point 105, 91 and 7 are both in these two graphs, which means these points have the large influential effect on the line. These points' height and FEV1 are (105): 66, 460; (91): 68, 426; (7): 68, 425.

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1.80338E-07	20	0.0506167	33
2.86084E-07	139	0.0553527	7
3.77669E-07	144	0.0565644	91
5.30699E-07	34	0.0612574	105
8.03993E-07	38	0.0915288	90

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-4.38E-01	90	0.212038	65
-3.21E-01	33	0.245860	149
-2.33E-01	40	0.336884	7
-2.14E-01	50	0.340671	91
-2.14E-01	12	0.363051	105

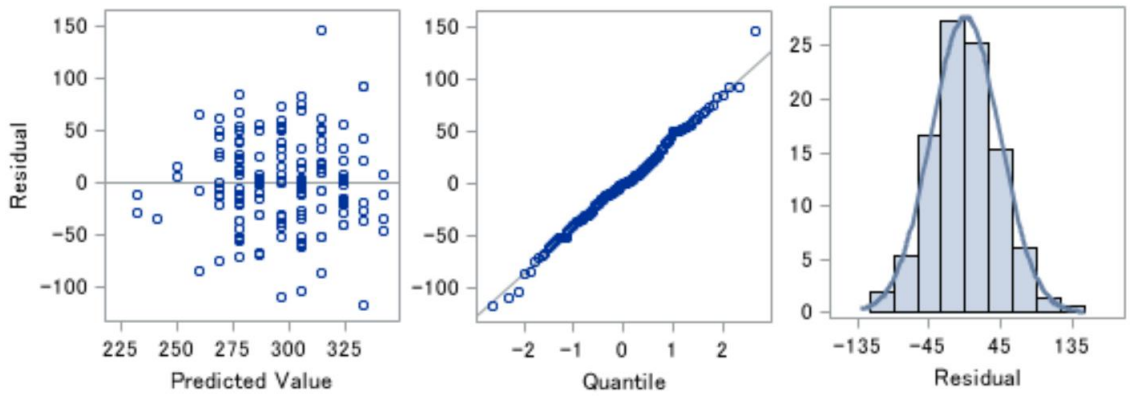
(6) Are the residuals normally distributed? Why?

Yes. The residuals are normally distributed.

As we see in the following graph 1, these points are distributed nearby the central line.

As we see in the following graph 2, these points are distributed like a line.

As we see in the following graph 3, the residual are normally distributed.



Problem 2

a=1+75=**76**

b=5117.5591/1=**5117.5591**

c=9879.2413/75=**131.7232**

d=5117.5591/131.7232=**38.8509**

e=5117.5591/14997=**0.3412**

f=76+1=**77**

Problem 3

See attached sas program “program3”. The “result” table in “main” is the answer.

	row	col	val
1	1	1	330
2	1	2	400
3	2	1	250
4	2	2	300
5	3	1	420
6	3	2	500
7	4	1	500
8	4	2	600