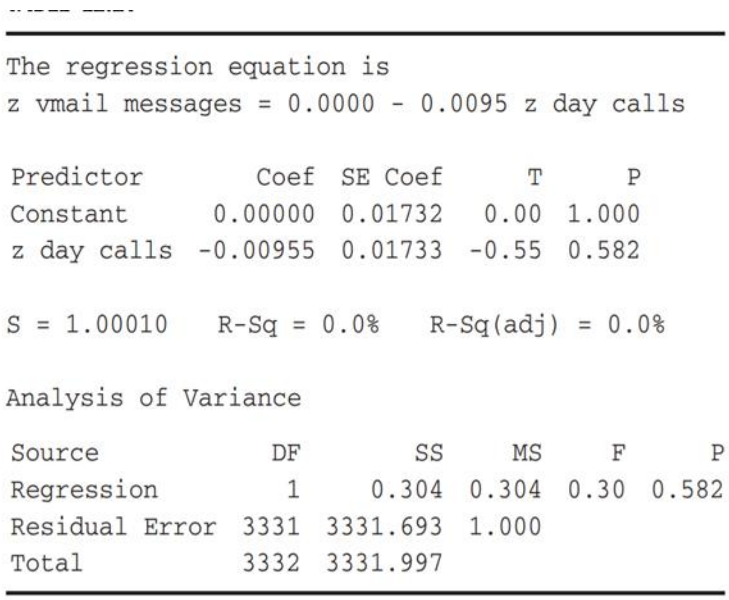
Bo Zhang 10411943 bzhang43@stevens.edu

#1 (15 points)

Based on the regression output shown in the Table below (from the churn data set), answer the following questions.

(a) Is there evidence of a linear relationship between z vmail messages (z-scores of the number of voice mail messages) and z day calls (z-scores of the number of day calls made)? Explain



```
The regression equation is
z vmail messages = 0.0000 - 0.0095 z day calls

Predictor        Coef  SE Coef      T      P
Constant      0.00000  0.01732   0.00  1.000
z day calls  -0.00955  0.01733  -0.55  0.582

S = 1.00010   R-Sq = 0.0%   R-Sq(adj) = 0.0%

Analysis of Variance

Source            DF        SS     MS     F      P
Regression         1     0.304  0.304  0.30  0.582
Residual Error  3331  3331.693  1.000
Total           3332  3331.997
```

Answer: There is not enough evidence of a linear relationship between z vmail messages and z day calls.
The reasons are as following:

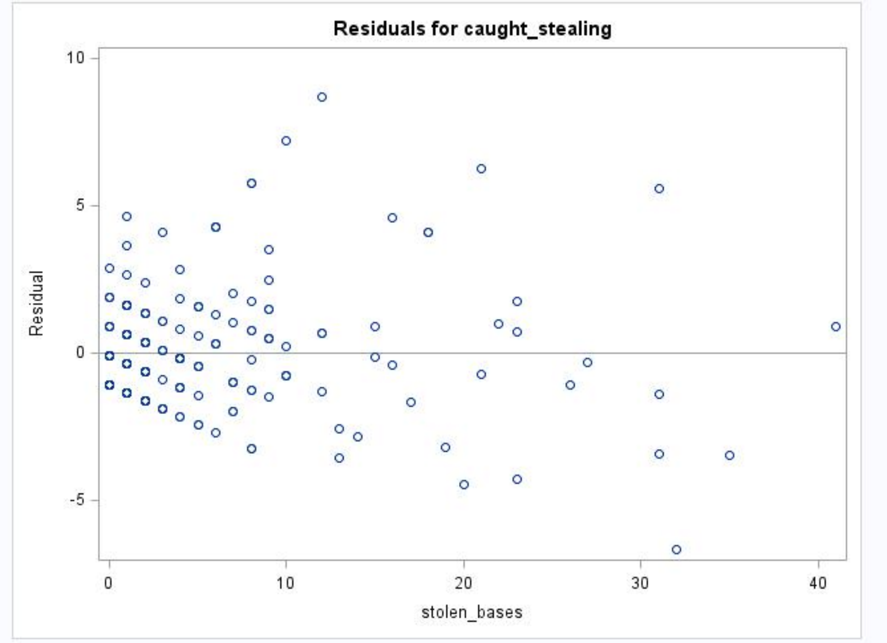
- For the model, p= 0.582>0.05, it means the F- test for the regression coefficient is not significant.
- For t-test of slope, p=0.582>0.05, and t-test for constant, p=1.000, the t-test results of both are not significant. And the result is the same with F-test.
- R2=0.0%, and adjusted R2=0.0%, this means it is hard for the linear regression to state the independent relationship among variables.

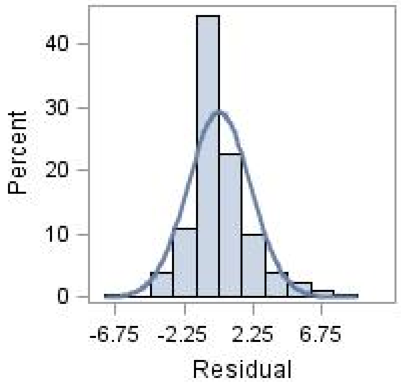
#2 (25 points)

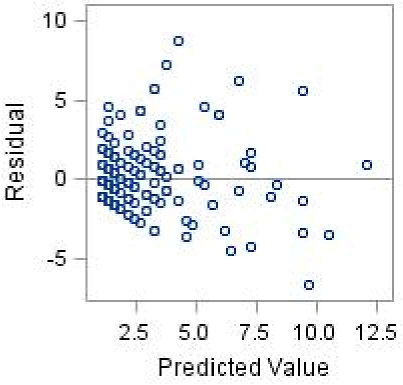
Open the baseball data set, which is available on the text book series website and CANVAS. Subset the data so that we are working with batters who have at least 100 at bats.

(a) We are interested in investigating whether there is a linear relationship between the number of times a player has been caught stealing and the number of stolen bases the player has. Construct a scatter plot with “caught” as the response. Is there evidence of a linear relationship?

Answer: The scatter plot indicates there may be a positive linear relationship between catch_stealing and stolen_bases. And a regression of caught_stealing on stolen_bases produced the normal probability plot of the standardized residuals, whose distribution is not normal, so there is no normality assumption. And the standardized residuals versus predicted values indicates nonconstant variable.





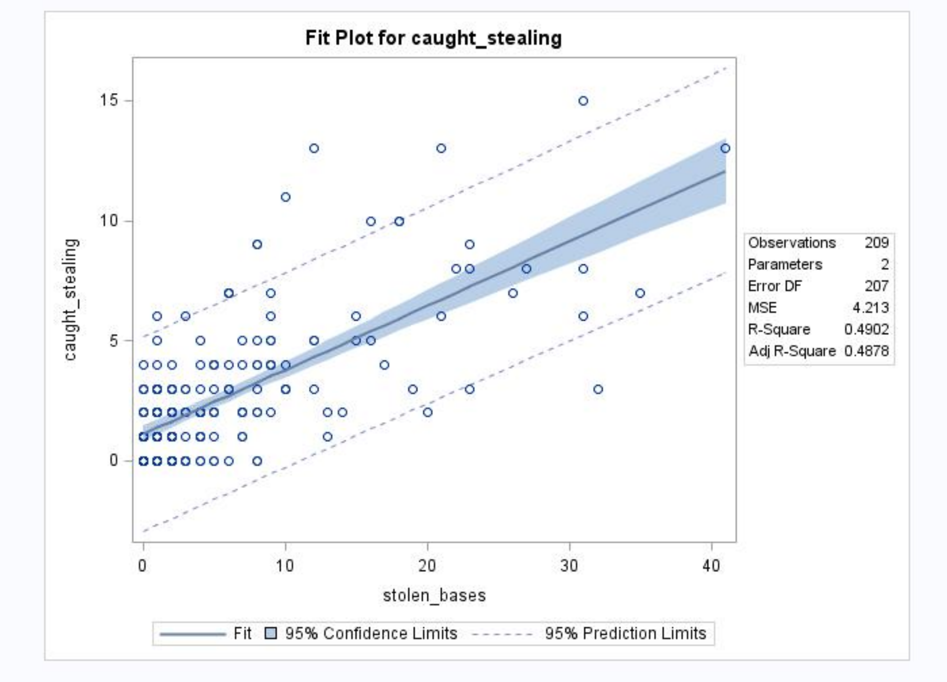


(b) Based on the scatter plot, is a transformation to linearity called for? Why or why not?

Answer: log transformation. Because if the relationship is not linear but curvilinear, it is not appropriate to model the relationship with a linear approximation. And log transformation may achieve linearity in the relationship.

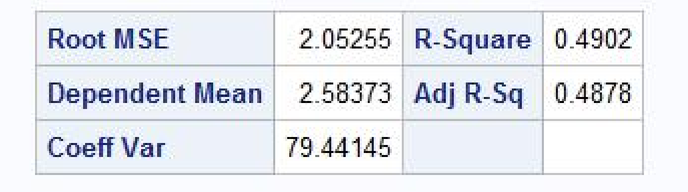
(c) Without any transformation, perform the regression of the number of times a player has been caught stealing versus the number of stolen bases the player has.

Answer: caught_stealing = 1.09863 + 0.26804*stolen_bases



(d) Find and interpret the statistic which tells you how well the data fit the model.

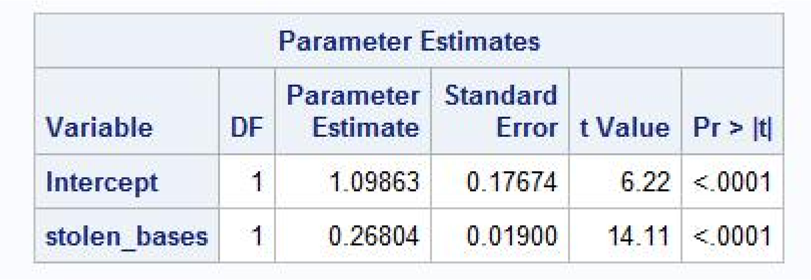
Answer: In general, 0<=R^2<=1, the higher the value is, the better the fit of the regression to the data set. In this model, R^2 =0.4902, so R=0.7001, it means the variables are positively correlated.



Root MSE	2.05255	R-Square	0.4902
Dependent Mean	2.58373	Adj R-Sq	0.4878
Coeff Var	79.44145		

(e) Interpret the y-intercept. Does this make sense? Why or why not?

Answer: y-intercept 1.09863 represents the estimated number of times a player has been caught with zero stolen bases the player has. It makes sense, because it meets the real world rules.



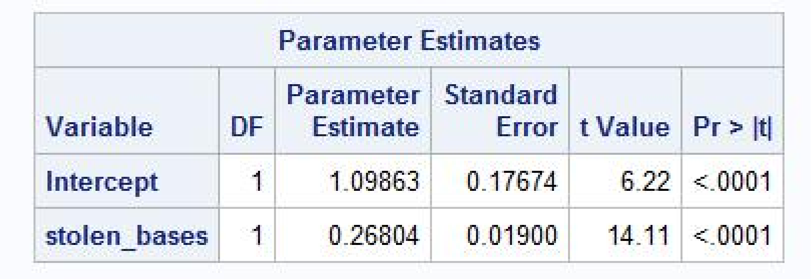
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.09863	0.17674	6.22	<.0001
stolen_bases	1	0.26804	0.01900	14.11	<.0001

(f) Inferentially, is there a significant relationship between the two variables? What tells you this?

Answer: Yes. There is a significant relationship between two variables.

Because from t-test, both regression coefficients are significant.

And the slope=0.26804, after calculating, 0 is not contained within confidence interval, so we can be sure of the significance of the relationship between the variables with 95% confidence.



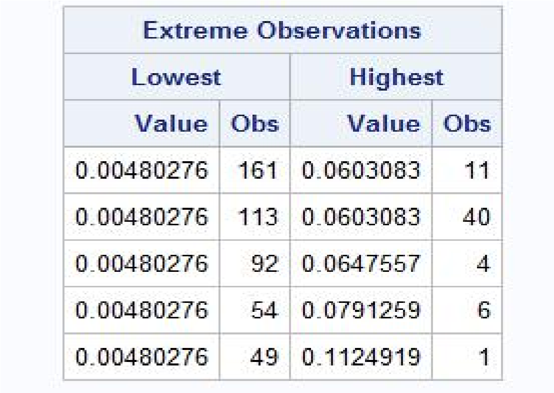
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.09863	0.17674	6.22	<.0001
stolen_bases	1	0.26804	0.01900	14.11	<.0001

(g) What are the influential observations?

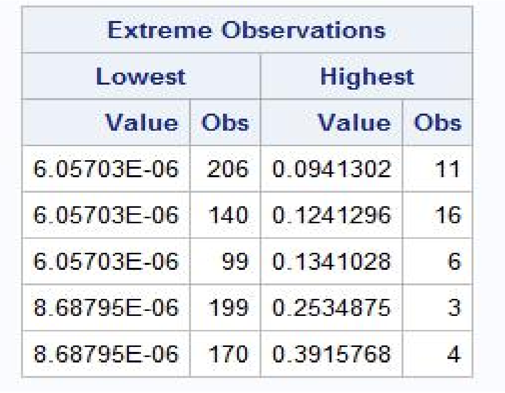
Answer: observation 4: Derek Jeter

observation 6: #Carlos Beltran

observation 11: *Johnny Damon



Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.00480276	161	0.0603083	11
0.00480276	113	0.0603083	40
0.00480276	92	0.0647557	4
0.00480276	54	0.0791259	6
0.00480276	49	0.1124919	1



Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
6.05703E-06	206	0.0941302	11
6.05703E-06	140	0.1241296	16
6.05703E-06	99	0.1341028	6
8.68795E-06	199	0.2534875	3
8.68795E-06	170	0.3915768	4

(h) What are the high leverage observations?

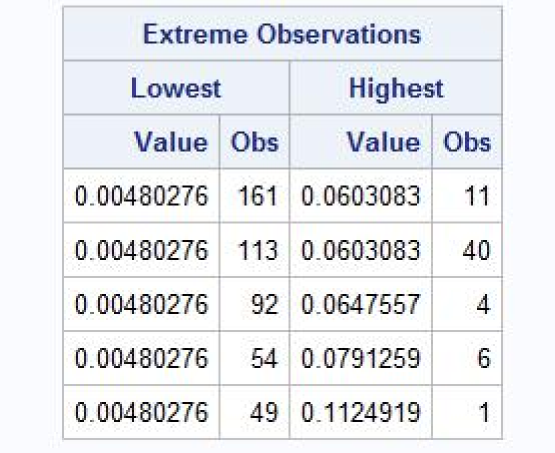
Answer: observation 1: Alfonso Sorian

observation 4: Derek Jeter

observation 6: #Carlos Beltran

observation 11: *Johnny Damon

observation 40: Mike Cameron



Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.00480276	161	0.0603083	11
0.00480276	113	0.0603083	40
0.00480276	92	0.0647557	4
0.00480276	54	0.0791259	6
0.00480276	49	0.1124919	1

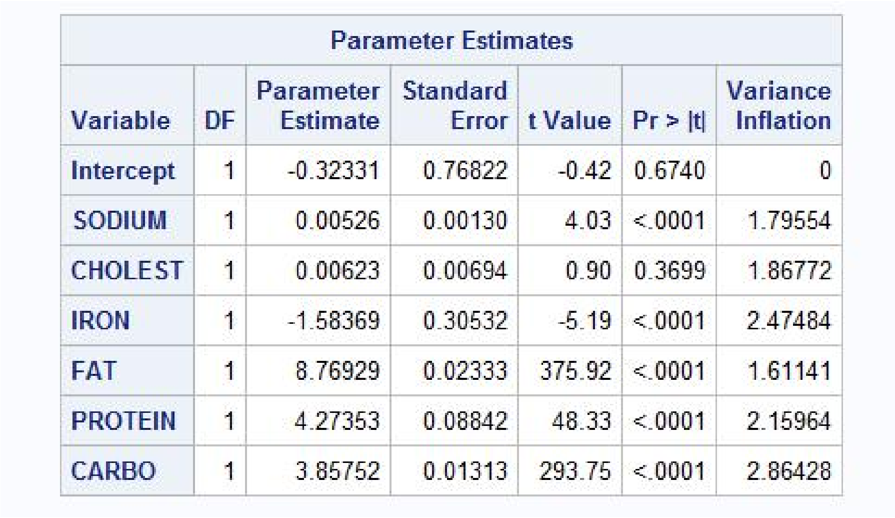
#3 (35 points)

Using the Nutrition data set on the text book series website and CANVAS, perform the follow analysis and answer the relevant questions.

(a) Create a regression model for dependent variable “calories” using predictors: sodium, cholesterol, iron, fat, protein, carbohydrates.

Answer:

calories=-0.32331+0.00526*sodium1.58369*iron+8.76929*fat+4.27353*protein+ 3.85752*carbo

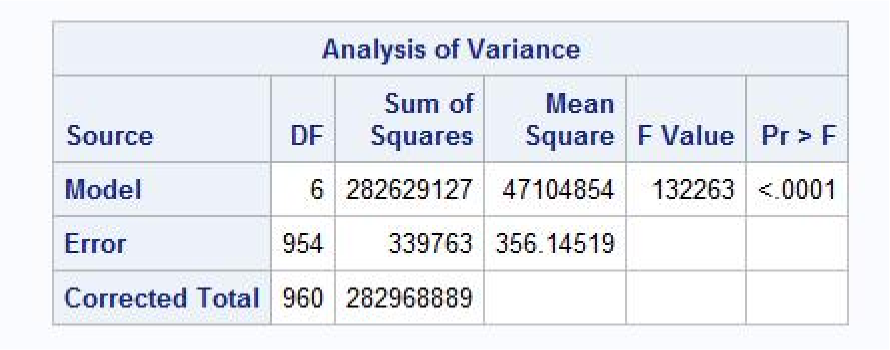


Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.32331	0.76822	-0.42	0.6740	0
SODIUM	1	0.00526	0.00130	4.03	<.0001	1.79554
CHOLEST	1	0.00623	0.00694	0.90	0.3699	1.86772
IRON	1	-1.58369	0.30532	-5.19	<.0001	2.47484
FAT	1	8.76929	0.02333	375.92	<.0001	1.61141
PROTEIN	1	4.27353	0.08842	48.33	<.0001	2.15964
CARBO	1	3.85752	0.01313	293.75	<.0001	2.86428

(b) What is the conclusion regarding the significance of the overall regression? How do you know? Does this mean that all the predictors are important? Explain.

Answer: Regarding the overall regression, it is significant.

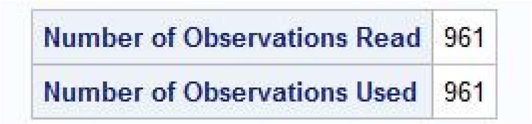
From the F-test, we know that p<0.0001<0.05, so the overall regression model is significant. But it does not mean all the predictors are important. Because F-test considers the linear relationship between the target variable and the set of predictors taken as a whole.



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	282629127	47104854	132263	<.0001
Error	954	339763	356.14519		
Corrected Total	960	282968889			

(c) How many foods are included in the sample?

Answer: There are 961.



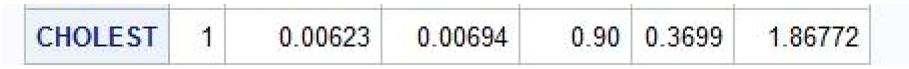
Number of Observations Read	961
Number of Observations Used	961

(d) How are we to interpret the value of b0, the coefficient for the constant term? Is this coefficient significantly different from zero? Explain how this makes sense.

Answer: For Intercept, it represents the estimated calories when all predictor variables equal zero. This coefficient equals -0.32331, it is not significantly different from zero. Because it is a model, it is not the value, it is estimated value.

(e) Which of the predictors probably does not belong in the model? Explain how you know this

Answer:It is cholest, because the p=0.3699, meaning it is not significant.



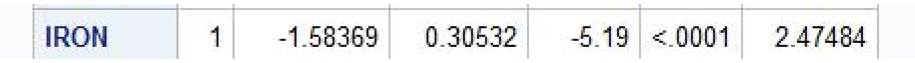
CHOLEST	1	0.00623	0.00694	0.90	0.3699	1.86772

(f) Suppose that we omit cholesterol from the model and rerun the regression. Explain what will happen to the value of R2.

Answer: R^2 may barely decrease, because when a predictor is removed from the model, the value of R^2 always goes down. If the predictor is useful, the value of R^2 will decrease significant; if the predictor is not useful, the value of R2 may barely decrease at all. Cholesterol is not significant, so R^2 may barely decrease, it is still 0.9988.

(g) Which predictor is negatively associated with the response? Explain how you know this.

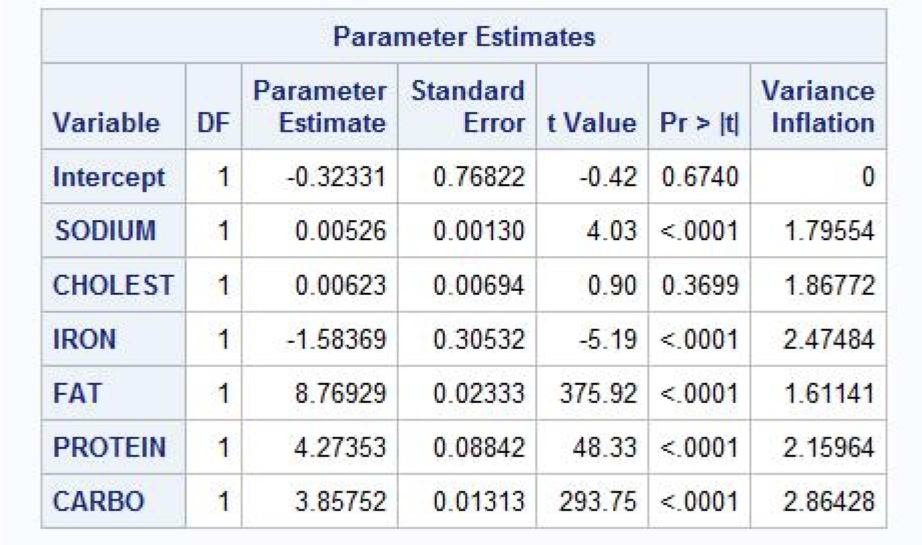
Answer: It is iron, because the coefficient -1.58369, is negative, indicating a negative relationship.



IRON	1	-1.58369	0.30532	-5.19	<.0001	2.47484

(h) Discuss the presence of multicollinearity. Evaluate the strength of evidence for the presence of multicollinearity.

Answer: The vif value for all the predictors are range from 1.6 to 2.9, indicating weak multicollinearity.

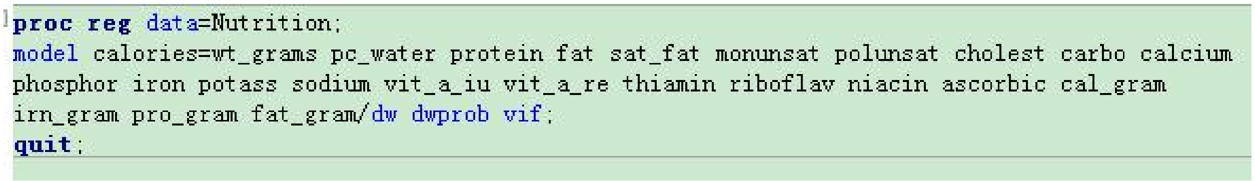


Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.32331	0.76822	-0.42	0.6740	0
SODIUM	1	0.00526	0.00130	4.03	<.0001	1.79554
CHOLEST	1	0.00623	0.00694	0.90	0.3699	1.86772
IRON	1	-1.58369	0.30532	-5.19	<.0001	2.47484
FAT	1	8.76929	0.02333	375.92	<.0001	1.61141
PROTEIN	1	4.27353	0.08842	48.33	<.0001	2.15964
CARBO	1	3.85752	0.01313	293.75	<.0001	2.86428

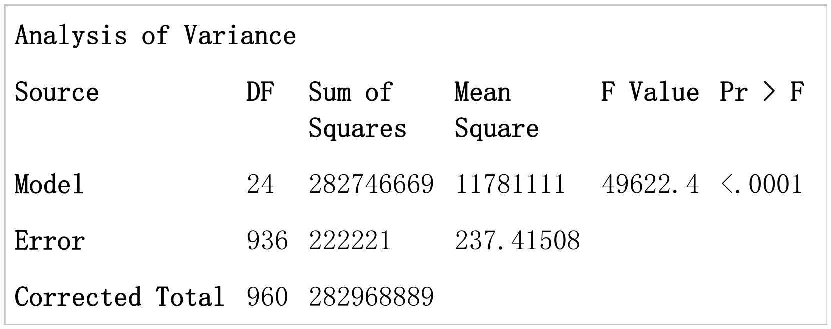
#4 (25 points)

Based on the Nutrition data set on the text book series website and CANVAS:

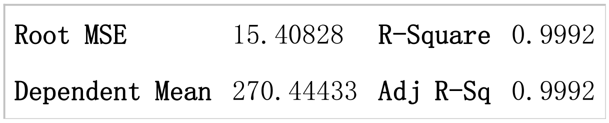
(a) Build the best multiple regression model you can for the purposes of predicting calories, using all the other variables as the predictors. Don't worry about whether or not the predictor coefficients are significant.



```
proc reg data=Nutrition;
model calories=wt_grams pc_water protein fat sat_fat monunsat polunsat cholest carbo calcium
phosphor iron potass sodium vit_a_iu vit_a_re thiamin riboflav niacin ascorbic cal_gram
irn_gram pro_gram fat_gram/dw dwprob vif;
quit;
```



```
Analysis of Variance

Source           DF   Sum of     Mean        F Value  Pr > F
                      Squares    Square

Model            24   282746669  11781111    49622.4  <.0001

Error            936  222221     237.41508

Corrected Total  960  282968889
```

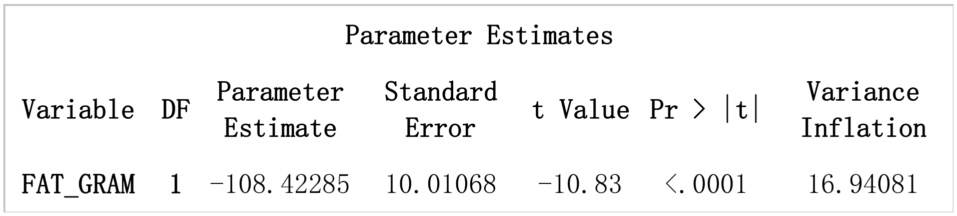


```
Root MSE        15.40828   R-Square  0.9992

Dependent Mean  270.44433  Adj R-Sq  0.9992
```

Coeff Var	5.69739

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-69.27787	6.93699	-9.99	<.0001	0
WT_GRAMS	1	0.03202	0.00842	3.80	0.0002	8.87614
PC_WATER	1	0.69184	0.07120	9.72	<.0001	20.99781
PROTEIN	1	4.47016	0.14758	30.29	<.0001	9.02486
FAT	1	10.50902	0.91174	11.53	<.0001	3692.67981
SAT_FAT	1	-2.00992	0.98818	-2.03	0.0422	455.82450
MONUNSAT	1	-1.82574	0.94886	-1.92	0.0546	719.41691
POLUNSAT	1	-1.66114	0.95438	-1.74	0.0821	490.15034
CHOLEST	1	0.01254	0.00709	1.77	0.0772	2.92121
CARBO	1	3.80651	0.01961	194.13	<.0001	9.58024
CALCIUM	1	0.02199	0.00628	3.50	0.0005	4.32969
PHOSPHOR	1	-0.02739	0.00657	-4.17	<.0001	7.30421
IRON	1	-2.48198	0.34959	-7.10	<.0001	4.86706
POTASS	1	-0.01963	0.00233	-8.42	<.0001	3.20860
SODIUM	1	0.00379	0.00113	3.35	0.0008	2.02901
VIT_A_IU	1	0.00028362	0.00035222	0.81	0.4209	7.45544
VIT_A_RE	1	-0.00163	0.00294	-0.55	0.5794	9.06512
THIAMIN	1	22.80544	3.70980	6.15	<.0001	5.25171
RIBOFLAV	1	1.74357	3.71071	0.47	0.6386	7.27788
NIACIN	1	0.30365	0.33633	0.90	0.3669	4.64541
ASCORBIC	1	-0.03652	0.01913	-1.91	0.0565	1.51073
CAL_GRAM	1	19.89873	1.82956	10.88	<.0001	50.71241
IRN_GRAM	1	33.50422	13.63249	2.46	0.0142	1.94629
PRO_GRAM	1	-20.03172	8.74247	-2.29	0.0222	2.49977



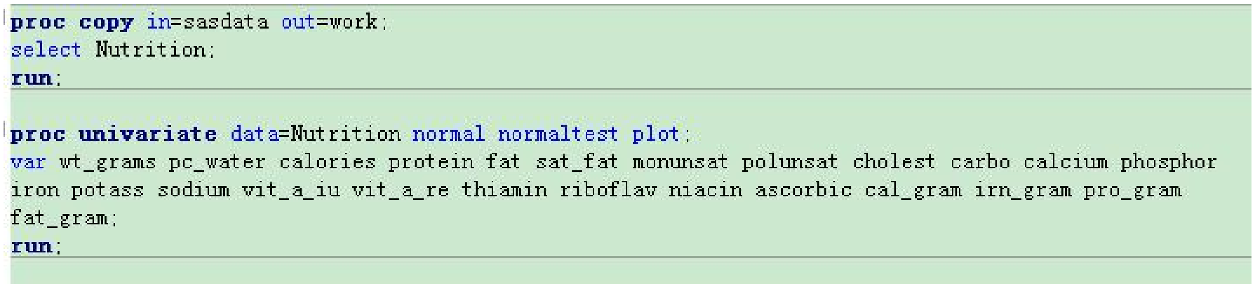
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
FAT_GRAM	1	-108.42285	10.01068	-10.83	<.0001	16.94081

So the model remove all predictors that p vlaues >0.05 and vif values are not acceptable.

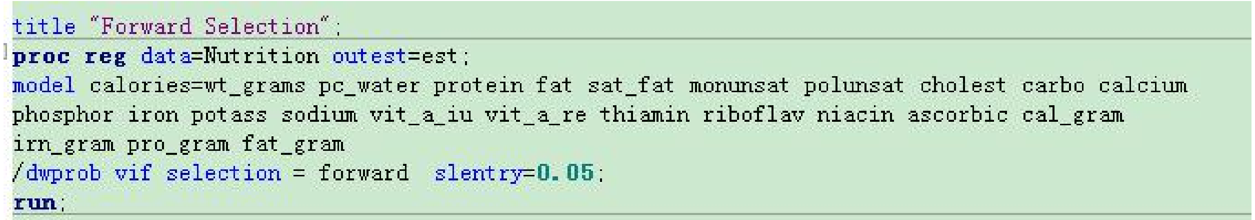
(b) Compare and contrast the results from the forward selection, backward elimination, and stepwise variable selection procedures.

Answer:



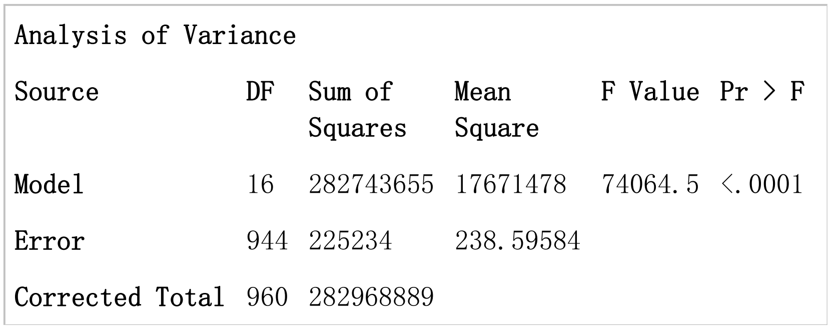
```
proc copy in=sasdata out=work;
select Nutrition;
run;

proc univariate data=Nutrition normal normaltest plot;
var wt_grams pc_water calories protein fat sat_fat monunsat polunsat cholest carbo calcium phosphor
iron potass sodium vit_a_iu vit_a_re thiamin riboflav niacin ascorbic cal_gram irn_gram pro_gram
fat_gram;
run;
```

qqplot of all predictors are not good.



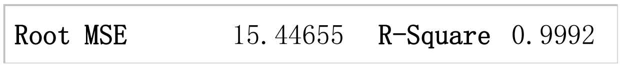
```
title "Forward Selection";
proc reg data=Nutrition outest=est;
model calories=wt_grams pc_water protein fat sat_fat monunsat polunsat cholest carbo calcium
phosphor iron potass sodium vit_a_iu vit_a_re thiamin riboflav niacin ascorbic cal_gram
irn_gram pro_gram fat_gram
/dwprob vif selection = forward  slentry=0.05;
run;
```

In forward selection, the model starts with no variables in it, and the variable with the highest sequential *F*-statistic is entered at each step.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	282743655	17671478	74064.5	<.0001
Error	944	225234	238.59584		
Corrected Total	960	282968889			



Root MSE	15.44655	R-Square	0.9992

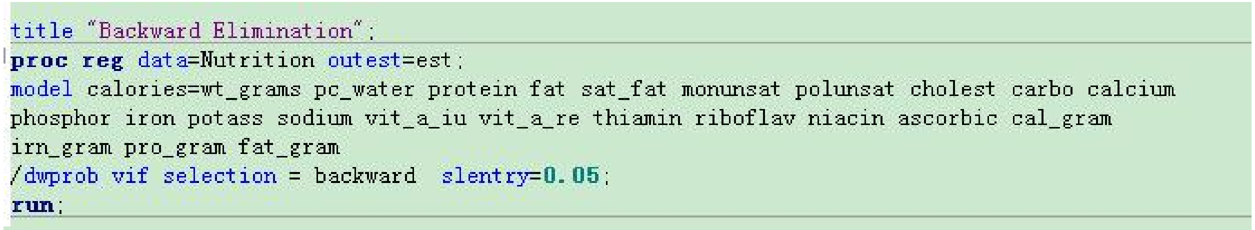
Dependent Mean	270.44433	Adj R-Sq	0.9992
Coeff Var	5.71154		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-68.99432	6.86782	-10.05	<.0001	0
WT_GRAMS	1	0.03072	0.00827	3.71	0.0002	8.51509
PC_WATER	1	0.68628	0.07054	9.73	<.0001	20.50411
PROTEIN	1	4.63655	0.12486	37.13	<.0001	6.42750
FAT	1	8.73727	0.03675	237.73	<.0001	5.97045
POLUNSAT	1	0.18041	0.07826	2.31	0.0214	3.27939
CARBO	1	3.81867	0.01858	205.57	<.0001	8.55574
CALCIUM	1	0.02020	0.00513	3.94	<.0001	2.87278
PHOSPHOR	1	-0.02677	0.00638	-4.20	<.0001	6.86016
IRON	1	-2.35077	0.34395	-6.83	<.0001	4.68812
POTASS	1	-0.02145	0.00202	-10.62	<.0001	2.39281
SODIUM	1	0.00392	0.00112	3.48	0.0005	1.99248
THIAMIN	1	23.58619	3.01033	7.84	<.0001	3.44090
CAL_GRAM	1	19.76562	1.80912	10.93	<.0001	49.34070
IRN_GRAM	1	35.92766	13.13549	2.74	0.0064	1.79802
PRO_GRAM	1	-20.46389	8.69956	-2.35	0.0189	2.46304
FAT_GRAM	1	-108.41735	9.87948	-10.97	<.0001	16.41802

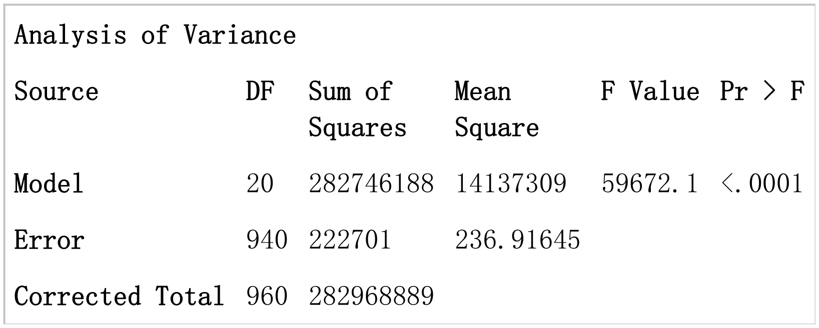
The model includes 16 variables in total. But residuals are not good.

and from vif, we know that pc_water, cal_gram and fat_gram are rather high.



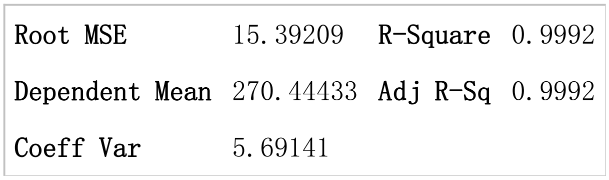
```
title “Backward Elimination”;
proc reg data=Nutrition outest=est;
model calories=wt_grams pc_water protein fat sat_fat monunsat polunsat cholest carbo calcium
phosphor iron potass sodium vit_a_iu vit_a_re thiamin riboflav niacin ascorbic cal_gram
irn_gram pro_gram fat_gram
/dwprob vif selection = backward  slentry=0.05;
run;
```

For the backward elimination procedure, the model begins with all of the variables in it, and the variable with the smallest partial *F*-statistic is removed.

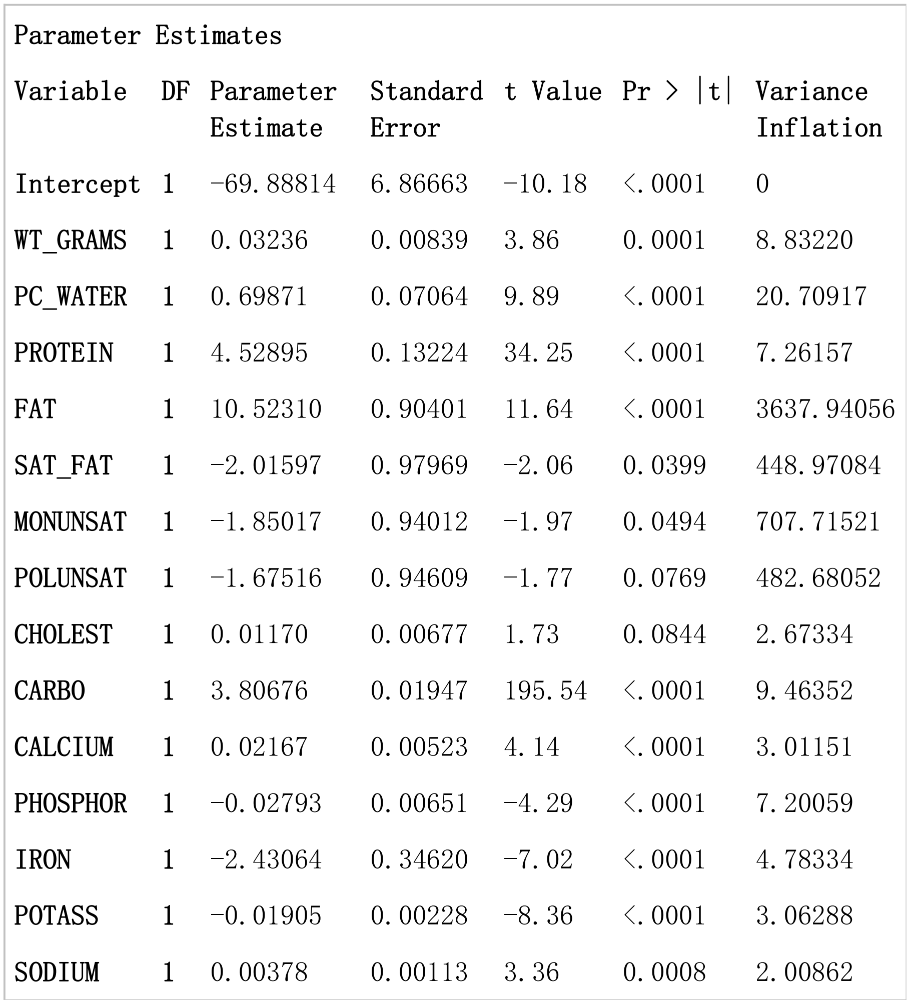


Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	282746188	14137309	59672.1	<.0001
Error	940	222701	236.91645		
Corrected Total	960	282968889			

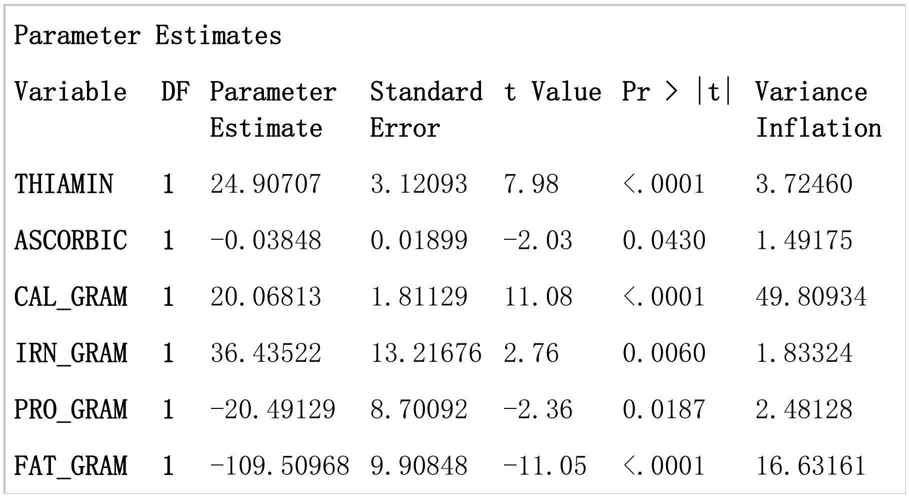


Root MSE	15.39209	R-Square	0.9992
Dependent Mean	270.44433	Adj R-Sq	0.9992
Coeff Var	5.69141		



Parameter Estimates

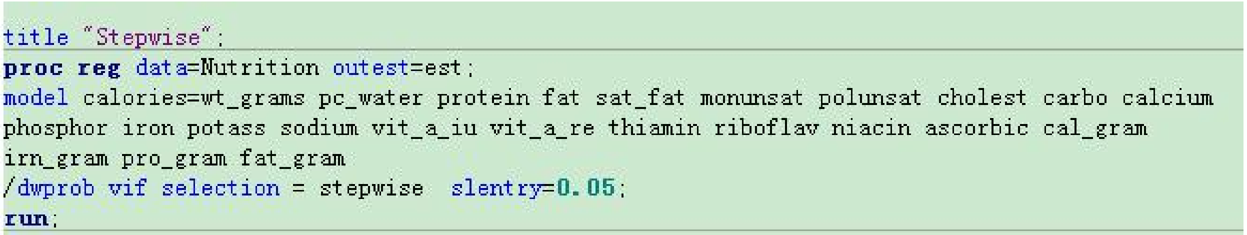
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-69.88814	6.86663	-10.18	<.0001	0
WT_GRAMS	1	0.03236	0.00839	3.86	0.0001	8.83220
PC_WATER	1	0.69871	0.07064	9.89	<.0001	20.70917
PROTEIN	1	4.52895	0.13224	34.25	<.0001	7.26157
FAT	1	10.52310	0.90401	11.64	<.0001	3637.94056
SAT_FAT	1	-2.01597	0.97969	-2.06	0.0399	448.97084
MONUNSAT	1	-1.85017	0.94012	-1.97	0.0494	707.71521
POLUNSAT	1	-1.67516	0.94609	-1.77	0.0769	482.68052
CHOLEST	1	0.01170	0.00677	1.73	0.0844	2.67334
CARBO	1	3.80676	0.01947	195.54	<.0001	9.46352
CALCIUM	1	0.02167	0.00523	4.14	<.0001	3.01151
PHOSPHOR	1	-0.02793	0.00651	-4.29	<.0001	7.20059
IRON	1	-2.43064	0.34620	-7.02	<.0001	4.78334
POTASS	1	-0.01905	0.00228	-8.36	<.0001	3.06288
SODIUM	1	0.00378	0.00113	3.36	0.0008	2.00862



Parameter Estimates

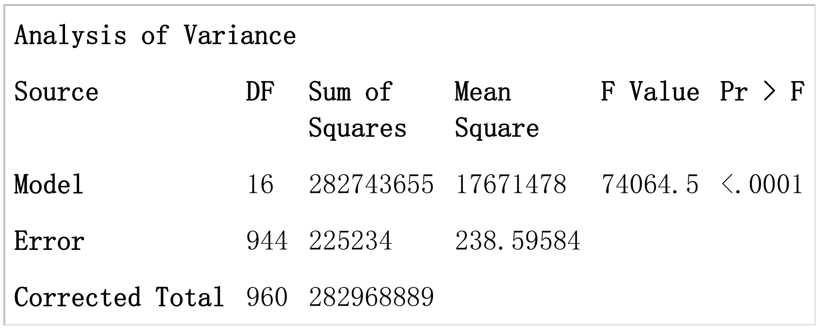
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
THIAMIN	1	24.90707	3.12093	7.98	<.0001	3.72460
ASCORBIC	1	-0.03848	0.01899	-2.03	0.0430	1.49175
CAL_GRAM	1	20.06813	1.81129	11.08	<.0001	49.80934
IRN_GRAM	1	36.43522	13.21676	2.76	0.0060	1.83324
PRO_GRAM	1	-20.49129	8.70092	-2.36	0.0187	2.48128
FAT_GRAM	1	-109.50968	9.90848	-11.05	<.0001	16.63161

The model includes 20 variables in total, and predictors riboflav, vit_a_re, vit_a_iu and niacin were removed from the model. But residuals are not good, and from vif, we know that pc_water, fat, sat_fat, monunsat, polunsat,cal_gram and fat_gram are rather high.



```
title “Stepwise”;
proc reg data=Nutrition outest=est;
model calories=wt_grams pc_water protein fat sat_fat monunsat polunsat cholest carbo calcium
phosphor iron potass sodium vit_a_iu vit_a_re thiamin riboflav niacin ascorbic cal_gram
irn_gram pro_gram fat_gram
/dwprob vif selection = stepwise  slentry=0.05;
run;
```

The stepwise modifies the forward selection procedure so that variables that have been entered into the model in earlier steps may still be withdrawn if they later turn out to be nonsignificant.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	282743655	17671478	74064.5	<.0001
Error	944	225234	238.59584		
Corrected Total	960	282968889			

Root MSE	15.44655	R-Square	0.9992
Dependent Mean	270.44433	Adj R-Sq	0.9992
Coeff Var	5.71154		

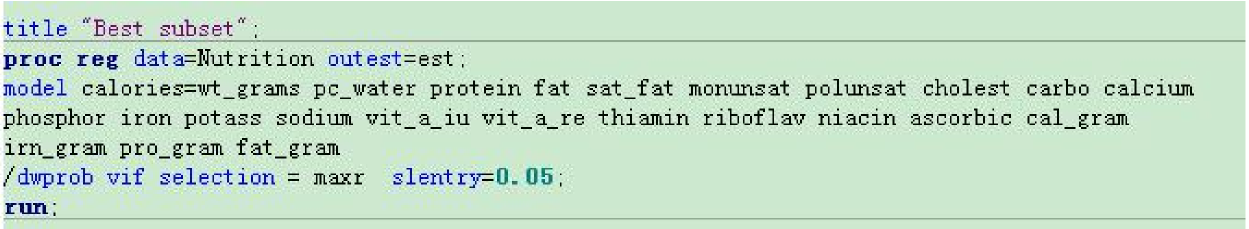
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-68.99432	6.86782	-10.05	<.0001	0
WT_GRAMS	1	0.03072	0.00827	3.71	0.0002	8.51509
PC_WATER	1	0.68628	0.07054	9.73	<.0001	20.50411
PROTEIN	1	4.63655	0.12486	37.13	<.0001	6.42750
FAT	1	8.73727	0.03675	237.73	<.0001	5.97045
POLUNSAT	1	0.18041	0.07826	2.31	0.0214	3.27939
CARBO	1	3.81867	0.01858	205.57	<.0001	8.55574
CALCIUM	1	0.02020	0.00513	3.94	<.0001	2.87278
PHOSPHOR	1	-0.02677	0.00638	-4.20	<.0001	6.86016
IRON	1	-2.35077	0.34395	-6.83	<.0001	4.68812
POTASS	1	-0.02145	0.00202	-10.62	<.0001	2.39281
SODIUM	1	0.00392	0.00112	3.48	0.0005	1.99248
THIAMIN	1	23.58619	3.01033	7.84	<.0001	3.44090
CAL_GRAM	1	19.76562	1.80912	10.93	<.0001	49.34070
IRN_GRAM	1	35.92766	13.13549	2.74	0.0064	1.79802
PRO_GRAM	1	-20.46389	8.69956	-2.35	0.0189	2.46304
FAT_GRAM	1	-108.41735	9.87948	-10.97	<.0001	16.41802

The model is the same as forward selection, includes 16 variables in total. But residuals are not good, and from vif, we know that pc_water, cal_gram and fat_gram are rather high.

In conclusion, forward selection and stepwise is more better.

(c) Apply the best subsets procedure, and compare against the previous methods.



```
title "Best subset";
proc reg data=Nutrition outest=est;
model calories=wt_grams pc_water protein fat sat_fat monunsat polunsat cholest carbo calcium
phosphor iron potass sodium vit_a_iu vit_a_re thiamin riboflav niacin ascorbic cal_gram
irn_gram pro_gram fat_gram
/dwprob vif selection = maxr  slentry=0.05;
run;
```

In the best subsets procedure, the software reports the best *k* models containing 1, 2, . . . , *p* predictors.

From all the subset model, this one has the most acceptable variables.

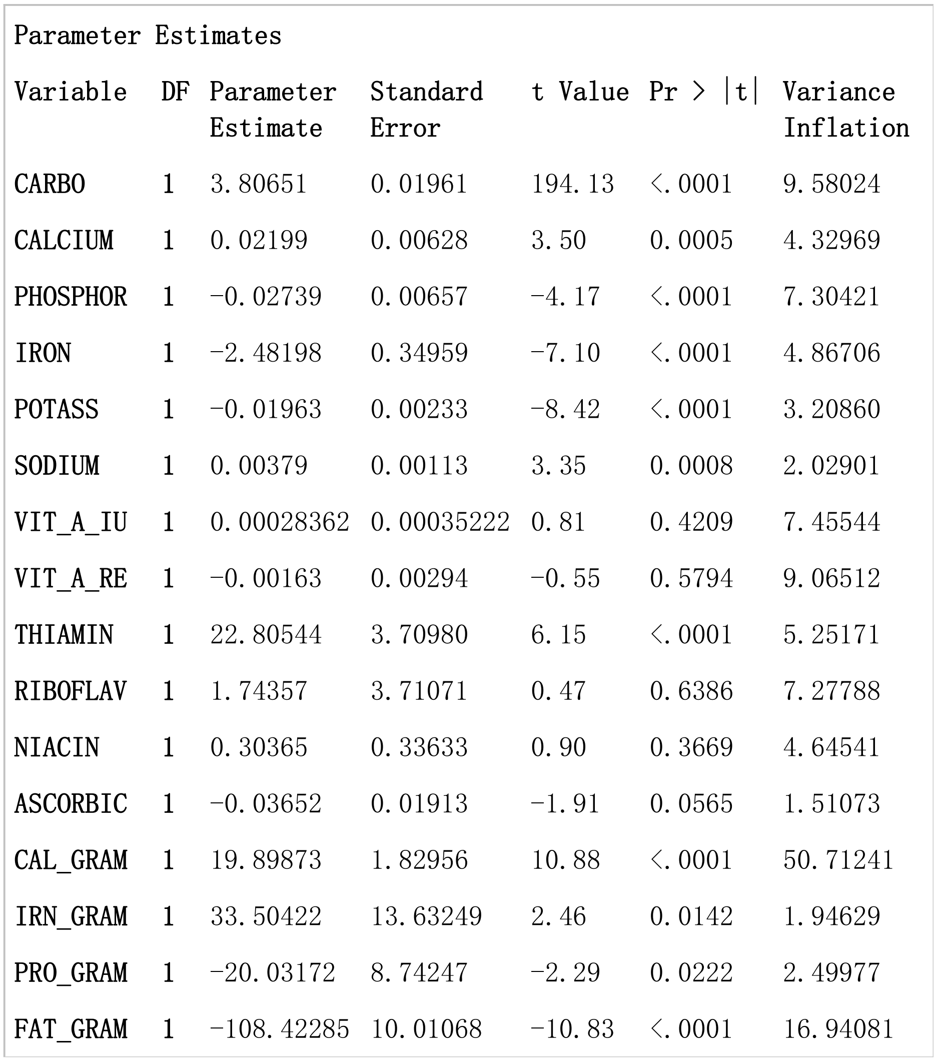
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	24	282746669	11781111	49622.4	<.0001
Error	936	222221	237.41508		
Corrected Total	960	282968889			

Root MSE	15.40828	R-Square	0.9992
Dependent Mean	270.44433	Adj R-Sq	0.9992
Coeff Var	5.69739		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-69.27787	6.93699	-9.99	<.0001	0
WT_GRAMS	1	0.03202	0.00842	3.80	0.0002	8.87614
PC_WATER	1	0.69184	0.07120	9.72	<.0001	20.99781
PROTEIN	1	4.47016	0.14758	30.29	<.0001	9.02486
FAT	1	10.50902	0.91174	11.53	<.0001	3692.67981
SAT_FAT	1	-2.00992	0.98818	-2.03	0.0422	455.82450
MONUNSAT	1	-1.82574	0.94886	-1.92	0.0546	719.41691
POLUNSAT	1	-1.66114	0.95438	-1.74	0.0821	490.15034
CHOLEST	1	0.01254	0.00709	1.77	0.0772	2.92121



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
CARBO	1	3.80651	0.01961	194.13	<.0001	9.58024
CALCIUM	1	0.02199	0.00628	3.50	0.0005	4.32969
PHOSPHOR	1	-0.02739	0.00657	-4.17	<.0001	7.30421
IRON	1	-2.48198	0.34959	-7.10	<.0001	4.86706
POTASS	1	-0.01963	0.00233	-8.42	<.0001	3.20860
SODIUM	1	0.00379	0.00113	3.35	0.0008	2.02901
VIT_A_IU	1	0.00028362	0.00035222	0.81	0.4209	7.45544
VIT_A_RE	1	-0.00163	0.00294	-0.55	0.5794	9.06512
THIAMIN	1	22.80544	3.70980	6.15	<.0001	5.25171
RIBOFLAV	1	1.74357	3.71071	0.47	0.6386	7.27788
NIACIN	1	0.30365	0.33633	0.90	0.3669	4.64541
ASCORBIC	1	-0.03652	0.01913	-1.91	0.0565	1.51073
CAL_GRAM	1	19.89873	1.82956	10.88	<.0001	50.71241
IRN_GRAM	1	33.50422	13.63249	2.46	0.0142	1.94629
PRO_GRAM	1	-20.03172	8.74247	-2.29	0.0222	2.49977
FAT_GRAM	1	-108.42285	10.01068	-10.83	<.0001	16.94081

The best subset model is the same as forward selection and stepwise procedure, includes 16 variables in total. But residuals are not good, and from vif, we know that pc_water, cal_gram and fat_gram are rather high.