1. 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 stable.

**proc** **copy** in = sasdata out = work;

select Nutrition;

**run**;

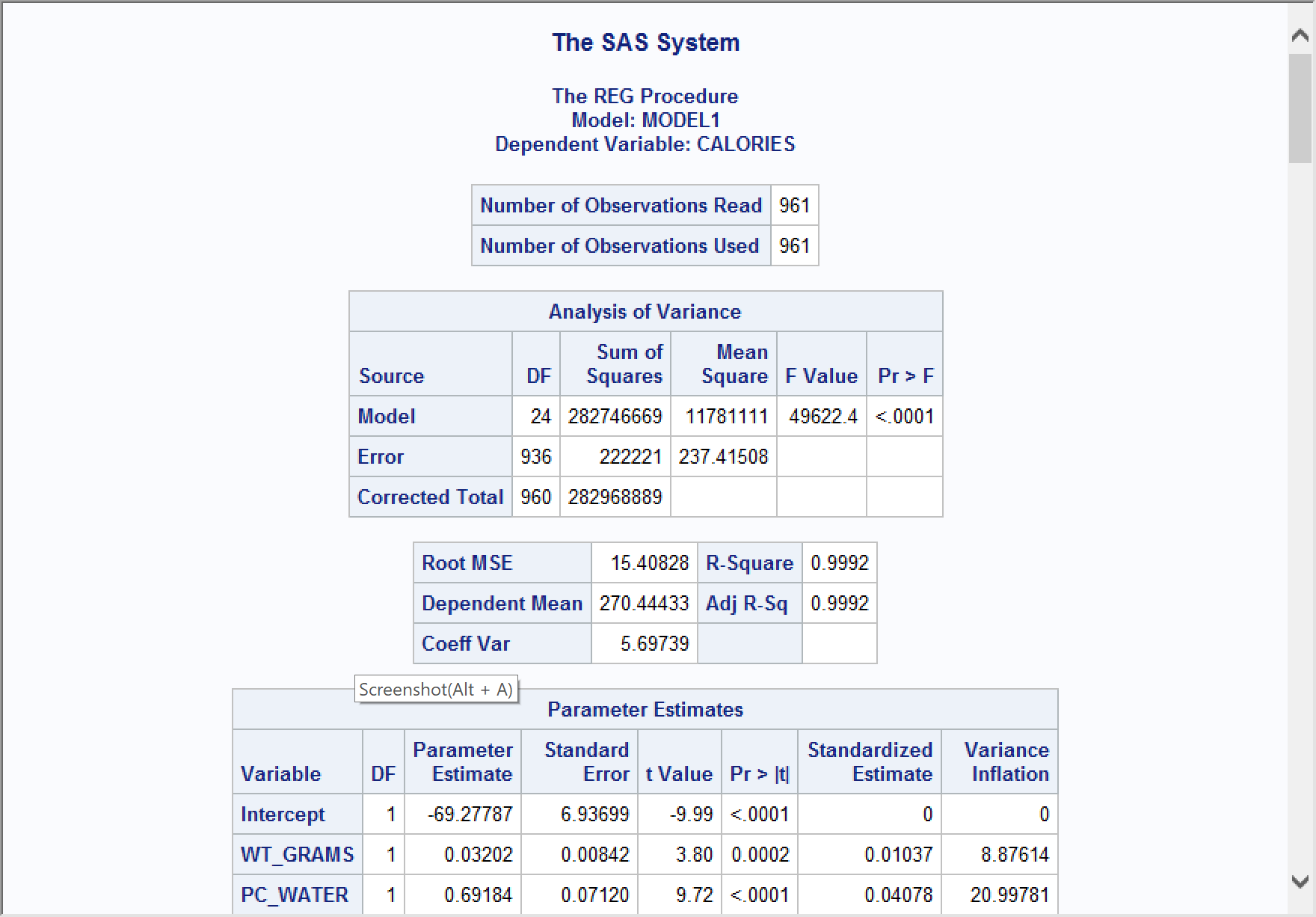
**proc** **reg** data = Nutrition;

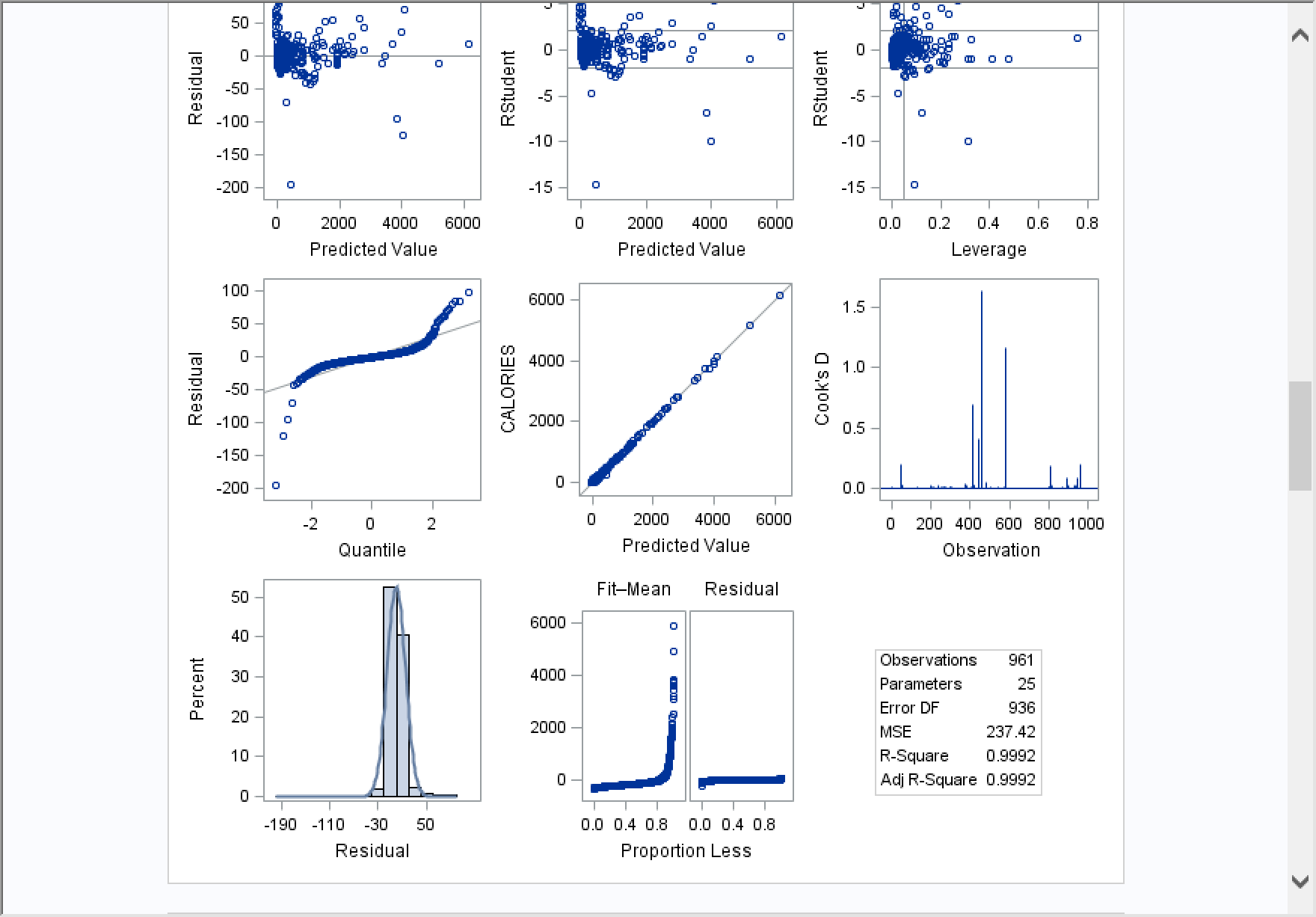
model CALORIES = WT\_GRAMS PC\_WATER PROTEIN FAT SAT\_FAT MONUNSAT POLUNSAT CHOLEST CARBO CALCIUM PHOSPHOR IRON POTASS SODIUM VIT\_A\_IU VIT\_A\_RE THIAMIN RIBOFLAV NIACIN ASCORBIC CAL\_GRAM IRN\_GRAM PRO\_GRAM FAT\_GRAM / stb vif dwProb dw;

OUTPUT OUT= regression

h= lev cookd= Cookd dffits= dffit;

**quit**;





1. Compare and contrast the results from the forward selection, backward elimination, and stepwise variable selection procedures.

Forward: Does not go back ,every step add the variables together.

/\*forward:\*/

**proc** **reg** data = Nutrition;

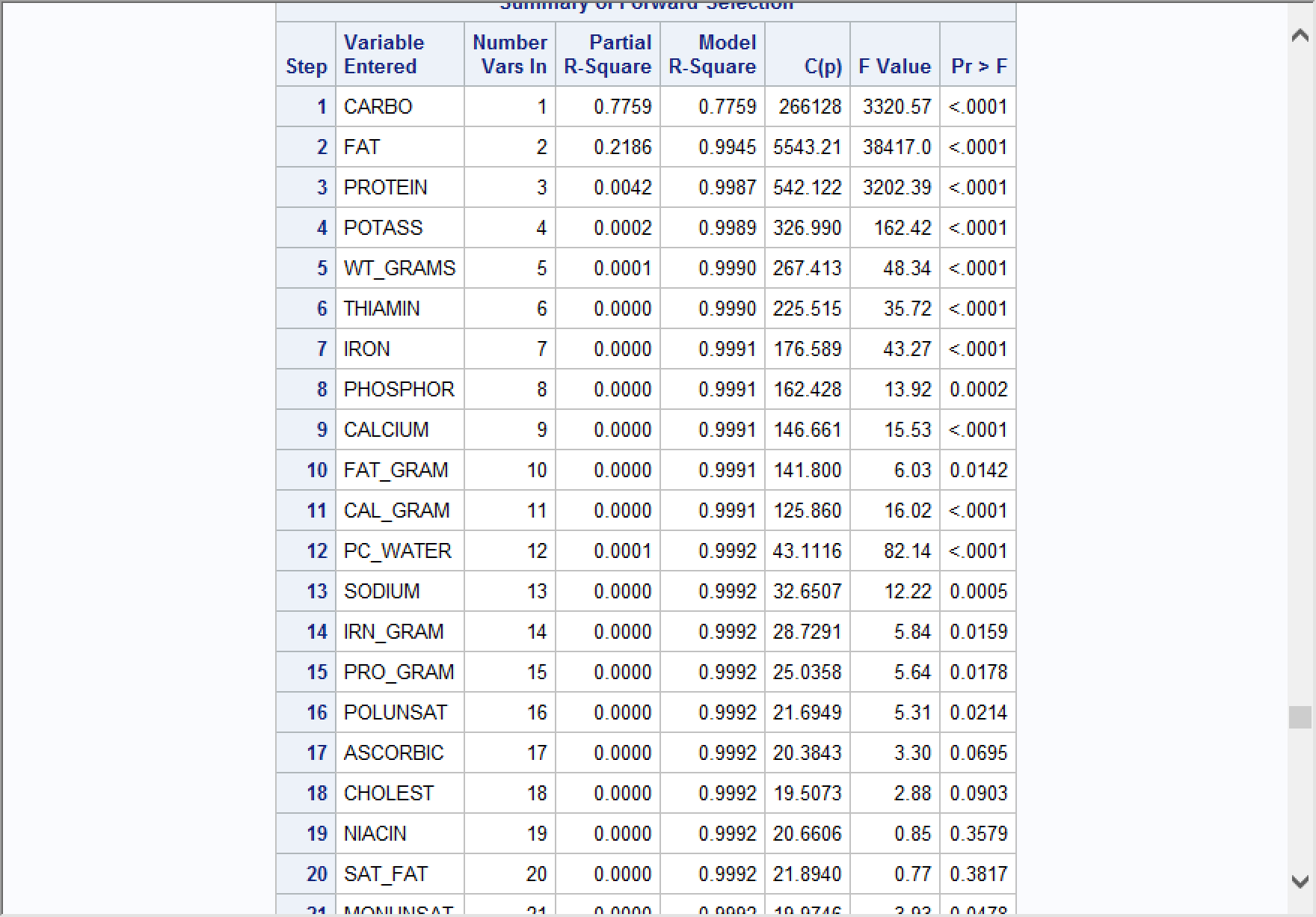
model CALORIES = WT\_GRAMS PC\_WATER PROTEIN FAT SAT\_FAT MONUNSAT POLUNSAT CHOLEST CARBO CALCIUM PHOSPHOR IRON POTASS SODIUM VIT\_A\_IU VIT\_A\_RE THIAMIN RIBOFLAV NIACIN ASCORBIC CAL\_GRAM IRN\_GRAM PRO\_GRAM FAT\_GRAM

/ selection = forward stb vif dwProb dw;

OUTPUT OUT=reg

h= lev cookd= Cookd dffits= dffit;

**quit**;



Depending which is least important .20 criterion leads to different results.

/\*backward:\*/

**proc** **reg** data = Nutrition;

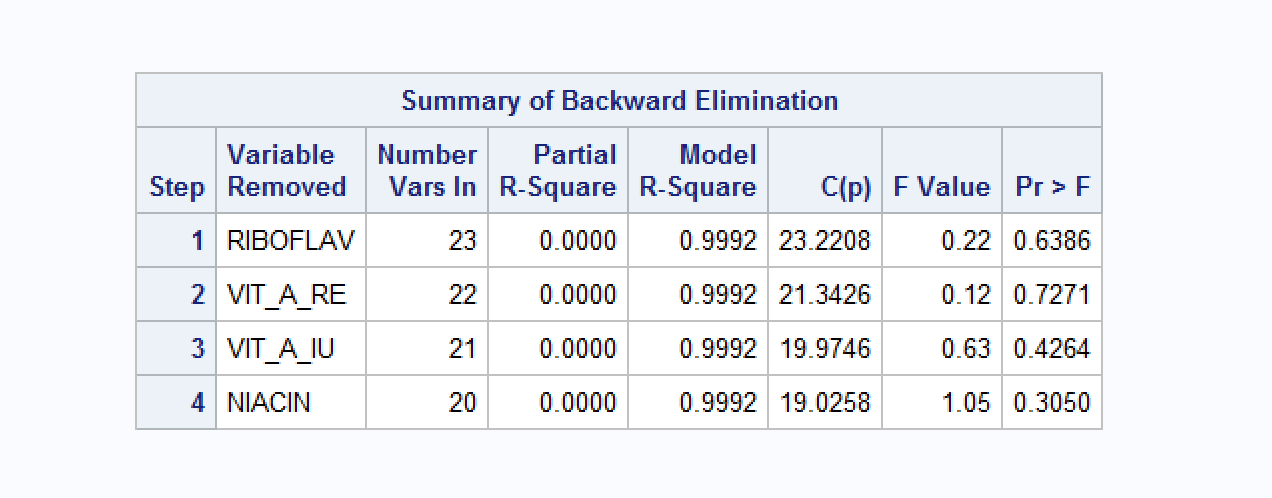
model CALORIES = WT\_GRAMS PC\_WATER PROTEIN FAT SAT\_FAT MONUNSAT POLUNSAT CHOLEST CARBO CALCIUM PHOSPHOR IRON POTASS SODIUM VIT\_A\_IU VIT\_A\_RE THIAMIN RIBOFLAV NIACIN ASCORBIC CAL\_GRAM IRN\_GRAM PRO\_GRAM FAT\_GRAM

/ selection = backward stb vif dwProb dw;

OUTPUT OUT=reg

h= lev cookd= Cookd dffits= dffit;

**quit**;



/\*stepwise\*/

**proc** **reg** data = Nutrition;

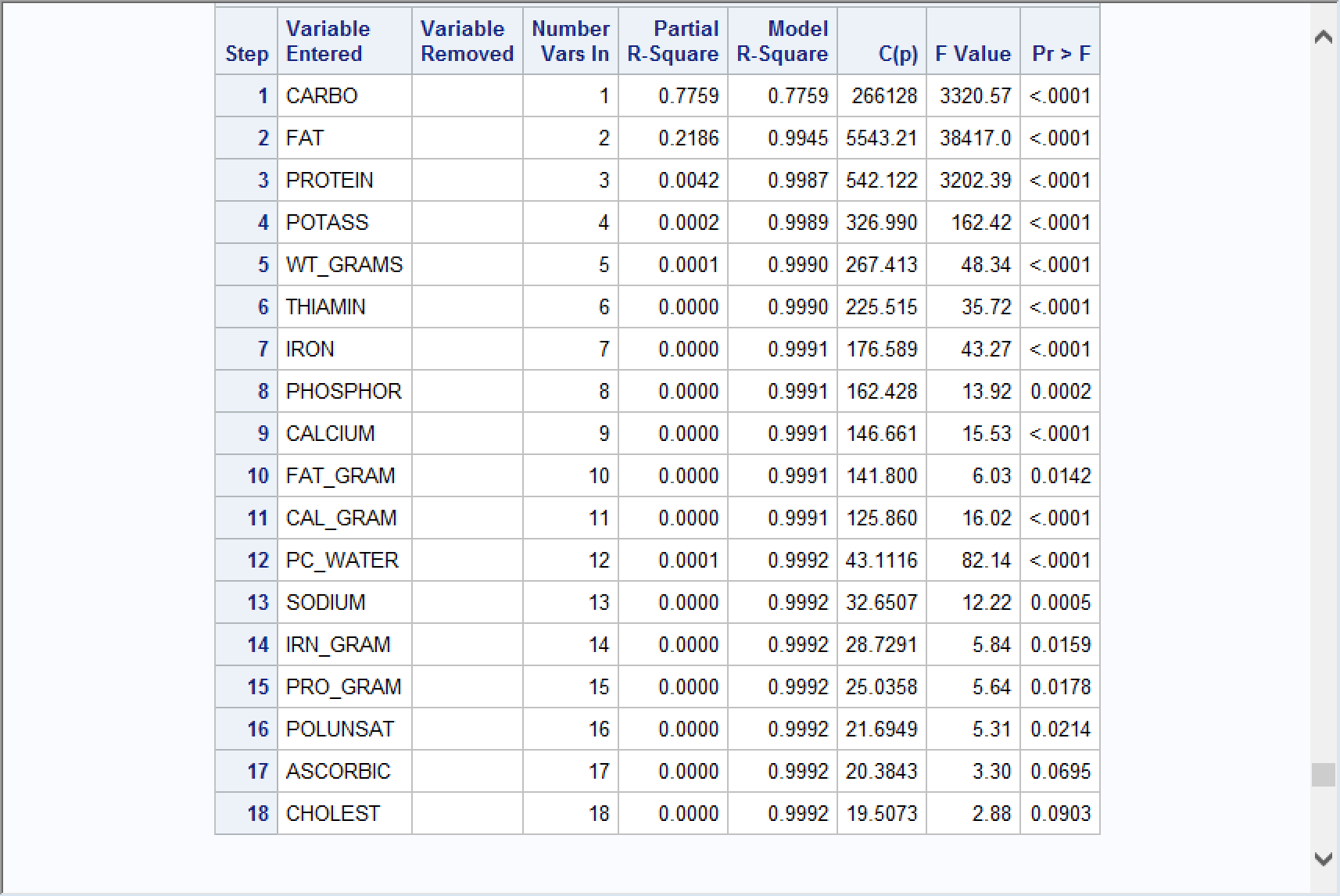
model CALORIES = WT\_GRAMS PC\_WATER PROTEIN FAT SAT\_FAT MONUNSAT POLUNSAT CHOLEST CARBO CALCIUM PHOSPHOR IRON POTASS SODIUM VIT\_A\_IU VIT\_A\_RE THIAMIN RIBOFLAV NIACIN ASCORBIC CAL\_GRAM IRN\_GRAM PRO\_GRAM FAT\_GRAM

/ selection = stepwise stb vif dwProb dw;

OUTPUT OUT=reg

h= lev cookd= Cookd dffits= dffit;

**quit**;



1. Apply the best subsets procedure, and compare against the previous methods.

Best solution if choose one variable , if choose 2 variable….etc

Give the top 5 important variables to predict.

If use pricinple components, they do not know what it is, but by using the best subsets procedure,we can know the relationships between all the variables.

/\*BEST\*/

**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

/ selection = MAXR stb vif dwProb dw;

OUTPUT OUT=reg

h= lev cookd= Cookd dffits= dffit;

**quit**;

