World happiness report 2021

Veronika Titarchuk

Group 8

05/01/2023

Stat 450-Sec02 Regression Analysis Spring 2023

Contents

Description of the data

The data describes how different factors such as GDP per capita, healthy life expectancy, social support, perception of corruption, generosity and freedom of life choices influence the happiness of people in each country. Using this data set, we can observe what is the most important factor for people to be happy, so that authorities can attempt to solve the issues that make people feel miserable. There are 149 observations, which represent each country. Overall, the World Happiness Report 2021 provides an important and timely contribution to our understanding of the complex and multifaceted nature of human happiness, and offers valuable insights for individuals, organizations, and policymakers seeking to promote greater well-being.

The data and the data-generating process

The happiness scores and rankings use data from the Gallup World Poll. Each regressor is crucial to understand human well-being. *Gross Domestic Product*, or how much each country produces, divided by the number of people in the country. GDP per capita gives information about the size of the economy and how the economy is performing.

Social support explains how having someone to count on in times of trouble influences our happiness.

Life expectancy describes how is your physical and mental health? Mental health is a key component of subjective well-being and is also a risk factor for future physical health and longevity. Mental health influences and drives a number of individual choices, behaviours, and outcomes.

Freedom to make life Choices describes a question “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”, which includes human rights, the right to life and liberty, freedom from slavery and torture, freedom of opinion and expression, the right to work and education, and many more.

Generosity is a certain indicator of a sense of active engagement in the community and a fundamental aspect of how people relate to one another.

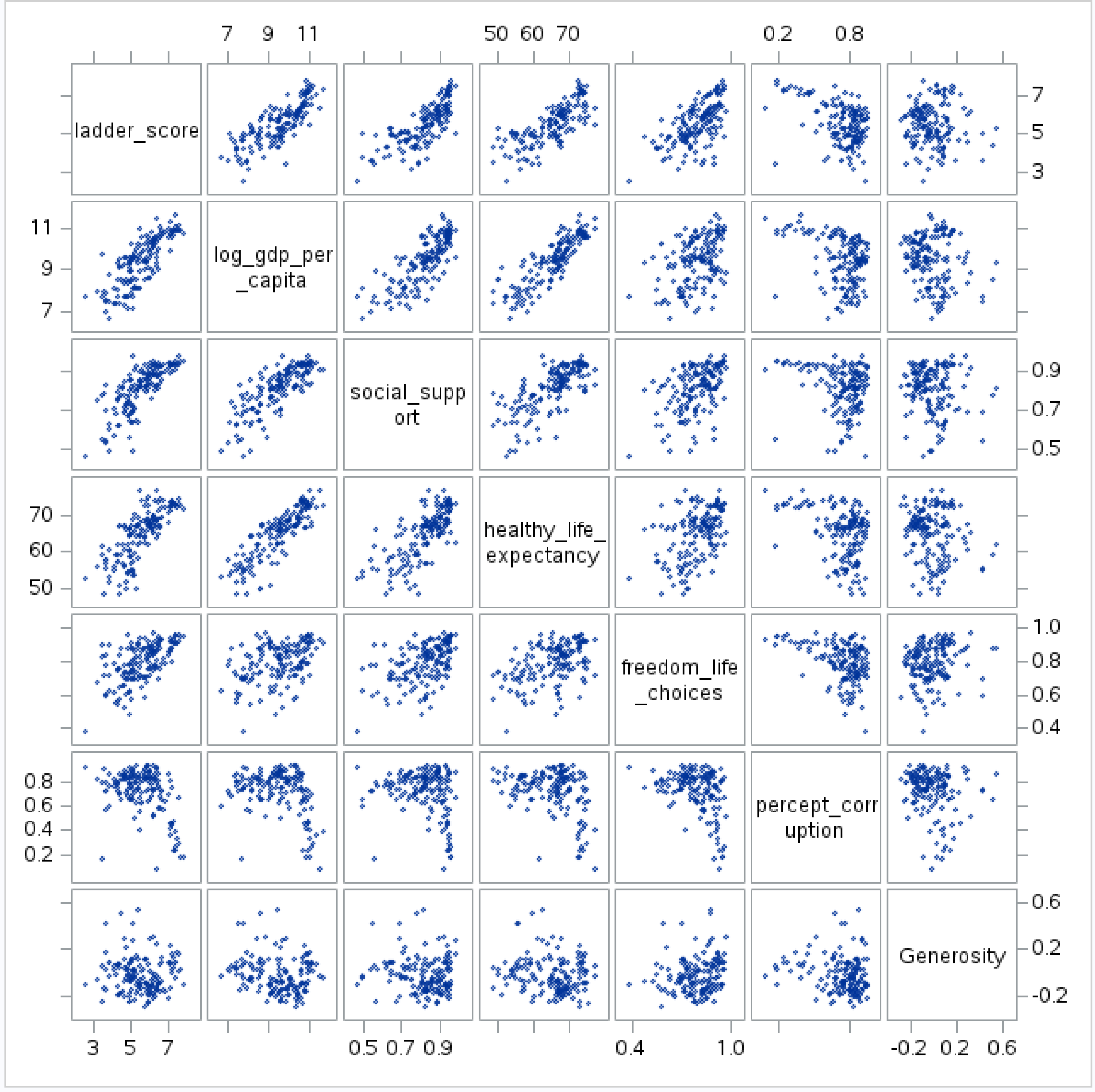
The perception of corruption explains whether or not individuals have faith in both the goodness of others and their own governments.

The World Happiness Report was written by a group of independent experts acting in their personal capacities. Any views expressed in this report do not necessarily reflect the views of any organization, agency or program of the United Nations.

The data didn't need to be cleaned since there are no empty values, but I had to drop columns such that std of ladder score, lower and upper whiskers, since they are variable that are parameters of the response, which in this data is the ladder score.

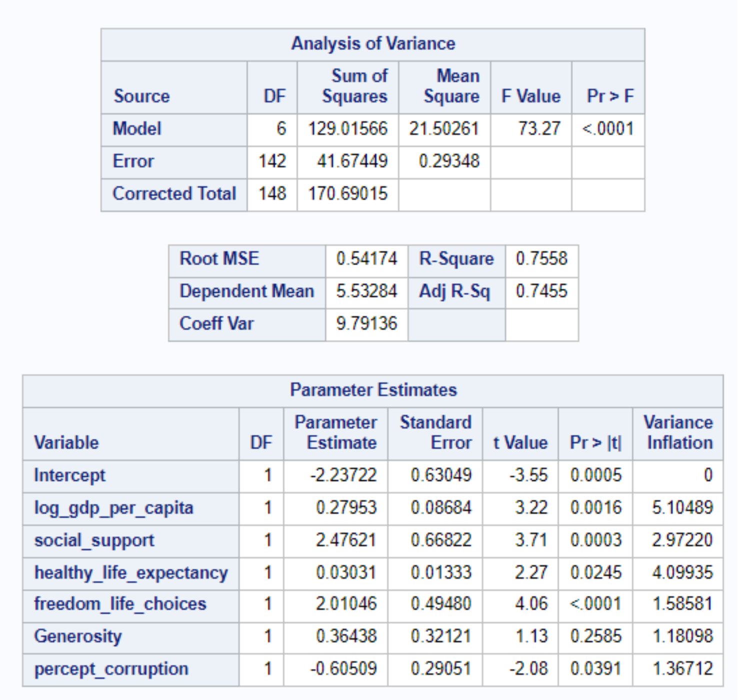
Exploratory data analysis

Selection of regressors



Looking at this scatter plot, we see that there are obviously some problems with linearity assumption in generosity and perception of corruption regressors.

Firstly, I will fit the linear regression to our model to see what problems we have and to see initial statistics, so we know later whether we are going in the right direction.

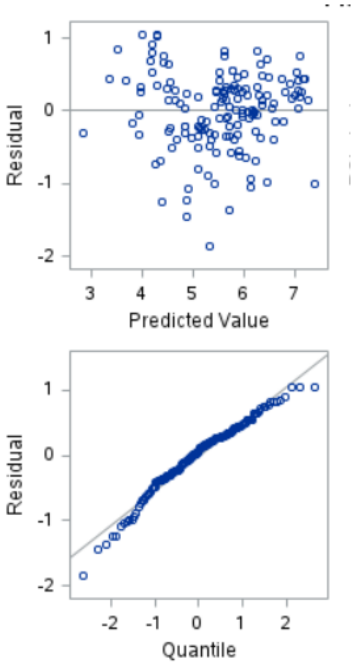


Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	129.01566	21.50261	73.27	<.0001
Error	142	41.67449	0.29348		
Corrected Total	148	170.69015			

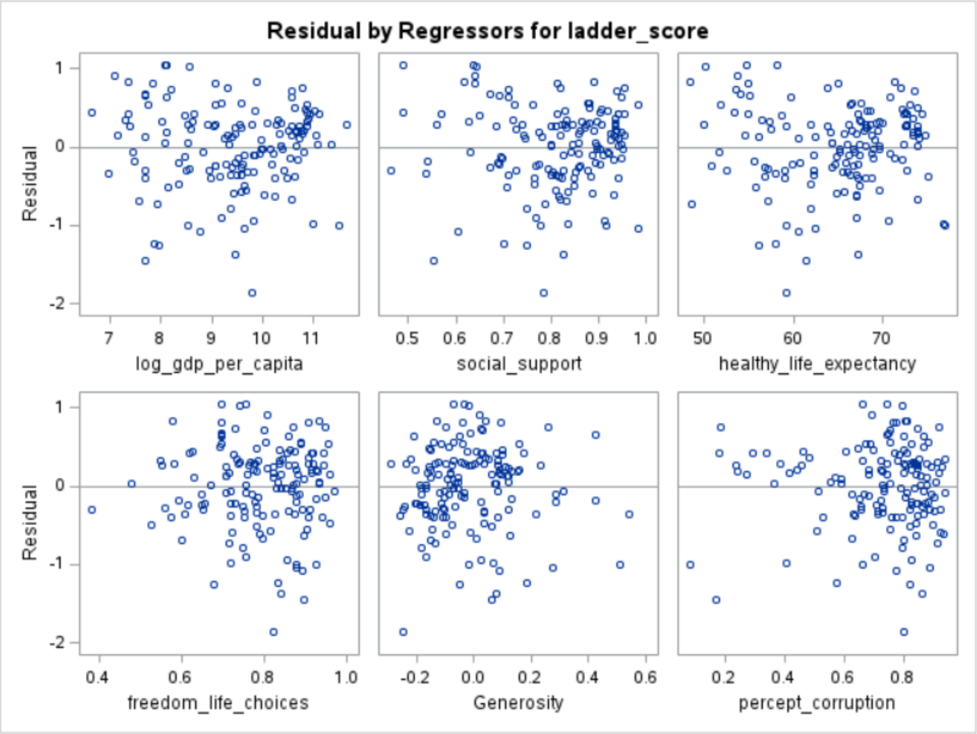
Root MSE	0.54174	R-Square	0.7558
Dependent Mean	5.53284	Adj R-Sq	0.7455
Coeff Var	9.79136		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-2.23722	0.63049	-3.55	0.0005	0
log_gdp_per_capita	1	0.27953	0.08684	3.22	0.0016	5.10489
social_support	1	2.47621	0.66822	3.71	0.0003	2.97220
healthy_life_expectancy	1	0.03031	0.01333	2.27	0.0245	4.09935
freedom_life_choices	1	2.01046	0.49480	4.06	<.0001	1.58581
Generosity	1	0.36438	0.32121	1.13	0.2585	1.18098
percept_corruption	1	-0.60509	0.29051	-2.08	0.0391	1.36712

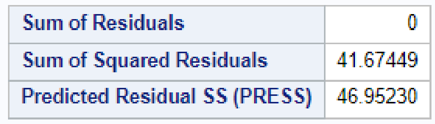
We see that we have a relatively high variance of inflation in GDP and healthy life expectancy model. R^2_{Adj} = 74.55% which is already good for 149 observations.



There are some problems with linearity assumption and constant variance assumption. We also have a little skewness of the data.

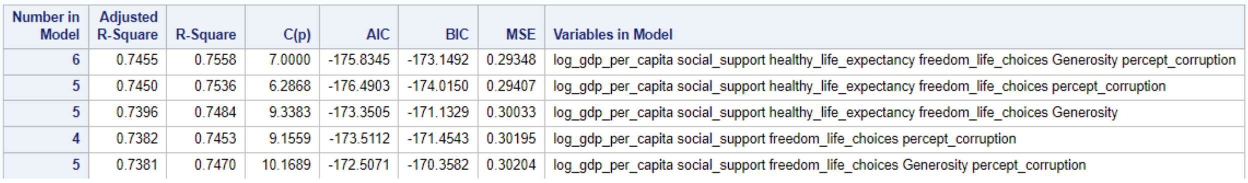


In residual by regressors plots, we see that we have some problems in generosity and perception of corruption variables.



Sum of Residuals	0
Sum of Squared Residuals	41.67449
Predicted Residual SS (PRESS)	46.95230

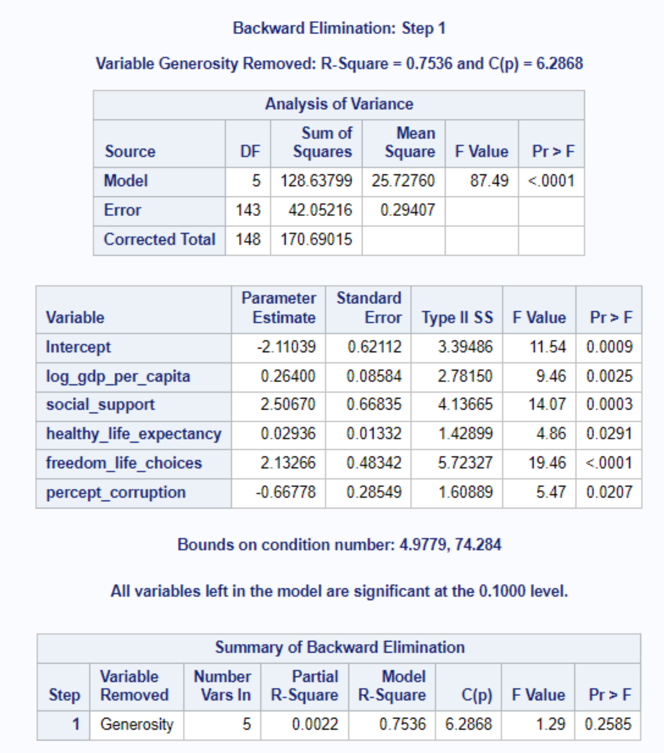
Using PRESS statistics, we can calculate the predictive capability of our model. R^2_{Pred} = 72.49%.



Number in Model	Adjusted R-Square	R-Square	C(p)	AIC	BIC	MSE	Variables in Model
6	0.7455	0.7558	7.0000	-175.8345	-173.1492	0.29348	log_gdp_per_capita social_support healthy_life_expectancy freedom_life_choices Generosity percept_corruption
5	0.7450	0.7536	6.2868	-176.4903	-174.0150	0.29407	log_gdp_per_capita social_support healthy_life_expectancy freedom_life_choices percept_corruption
5	0.7396	0.7484	9.3383	-173.3505	-171.1329	0.30033	log_gdp_per_capita social_support healthy_life_expectancy freedom_life_choices Generosity
4	0.7382	0.7453	9.1559	-173.5112	-171.4543	0.30195	log_gdp_per_capita social_support freedom_life_choices percept_corruption
5	0.7381	0.7470	10.1689	-172.5071	-170.3582	0.30204	log_gdp_per_capita social_support freedom_life_choices Generosity percept_corruption

Here we can see that AIC and BIC are pretty good and the data is unbiased. But from the regression we know that generosity variable is more likely insignificant. So, let’s try some selection techniques to see with which model we should work.

Backward elimination:



Backward Elimination: Step 1

Variable Generosity Removed: R-Square = 0.7536 and C(p) = 6.2868

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	128.63799	25.72760	87.49	<.0001
Error	143	42.05216	0.29407		
Corrected Total	148	170.69015			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-2.11039	0.62112	3.39486	11.54	0.0009
log_gdp_per_capita	0.26400	0.08584	2.78150	9.46	0.0025
social_support	2.50670	0.66835	4.13665	14.07	0.0003
healthy_life_expectancy	0.02936	0.01332	1.42899	4.86	0.0291
freedom_life_choices	2.13266	0.48342	5.72327	19.46	<.0001
percept_corruption	-0.66778	0.28549	1.60889	5.47	0.0207

Bounds on condition number: 4.9779, 74.284

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

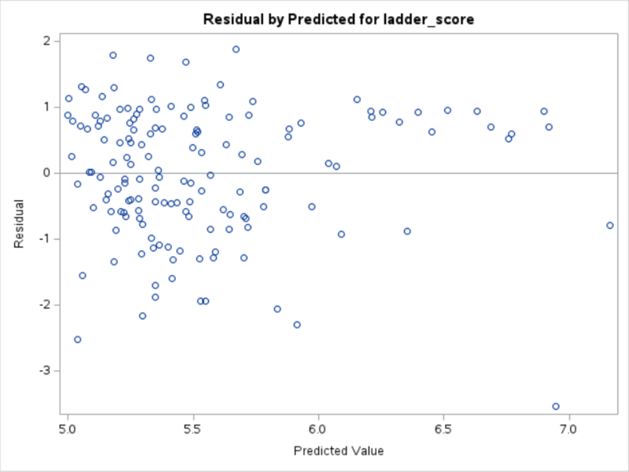
Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	Generosity	5	0.0022	0.7536	6.2868	1.29	0.2585

We see that the generosity variable dropped and now we have all the significant regressors at the 0.1 level. The stepwise selection showed the same result, and in forward selection generosity variable was added the last. So, we confirm that the generosity variable is not significant enough to include into our model.

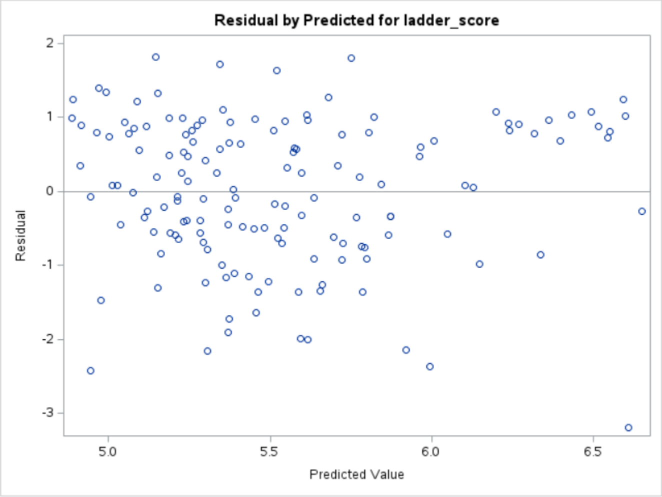
Since from the beginning we saw that we have problems with some variables, I will try to apply some transformations and see what changes.

Data Transformation

Perception of corruption initial residual vs predicted value plot:

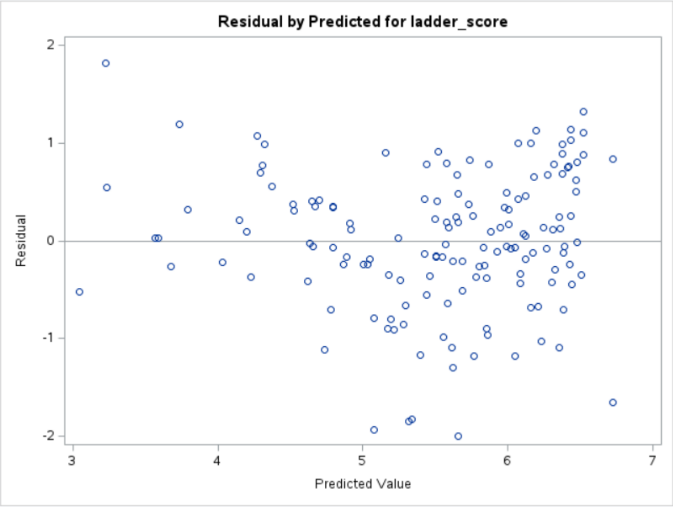


Perception of corruption squared residual vs predicted value plot:

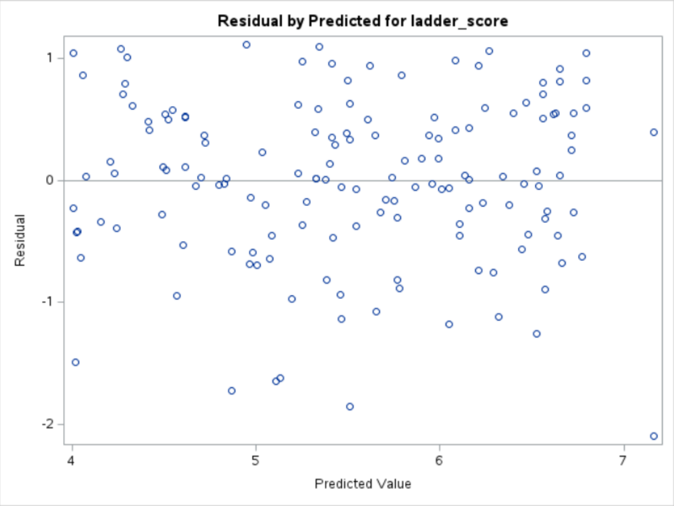


We see that we now have a better distribution.

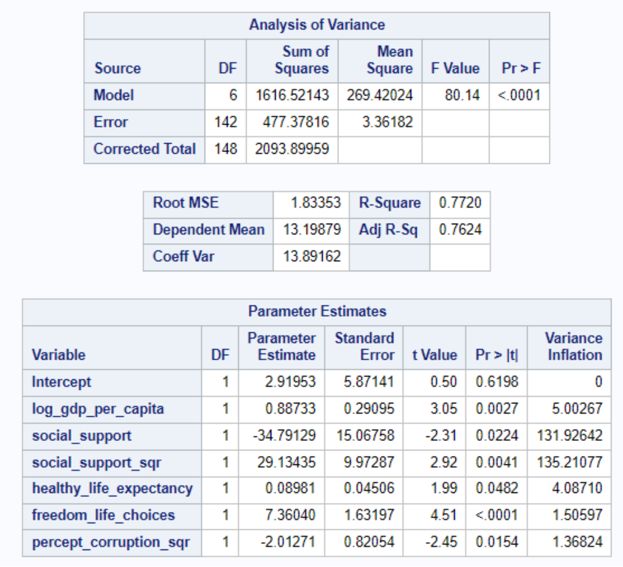
Social support initial residual vs predicted value plot:



Social support squared + social support residual vs predicted value plot:



The distribution got much better. Let’s fit the model using updated regressors.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1616.52143	269.42024	80.14	<.0001
Error	142	477.37816	3.36182		
Corrected Total	148	2093.89959			

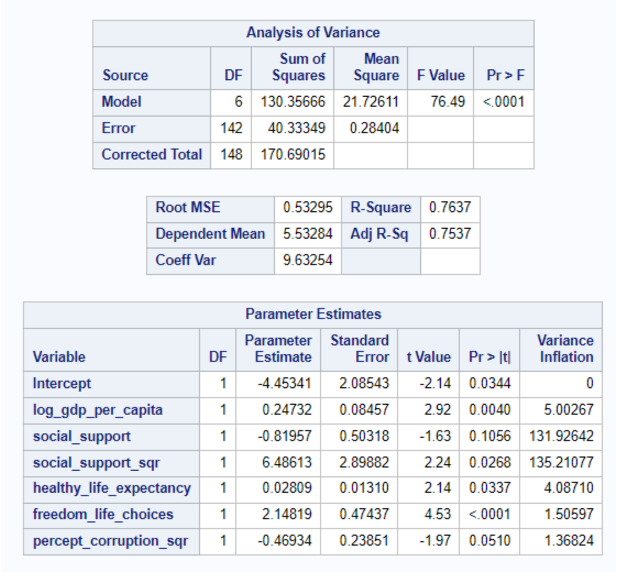
Root MSE	1.83353	R-Square	0.7720
Dependent Mean	13.19879	Adj R-Sq	0.7624
Coeff Var	13.89162		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	2.91953	5.87141	0.50	0.6198	0
log_gdp_per_capita	1	0.88733	0.29095	3.05	0.0027	5.00267
social_support	1	-34.79129	15.06758	-2.31	0.0224	131.92642
social_support_sqr	1	29.13435	9.97287	2.92	0.0041	135.21077
healthy_life_expectancy	1	0.08981	0.04506	1.99	0.0482	4.08710
freedom_life_choices	1	7.36040	1.63197	4.51	<.0001	1.50597
percept_corruption_sqr	1	-2.01271	0.82054	-2.45	0.0154	1.36824

We see that the model has better statistics, but now we see that social support and social support squared have huge VIFs.

Then, I was trying to normalize social support variable using proc stdize, and that’s the result I got:



Analysis of Variance

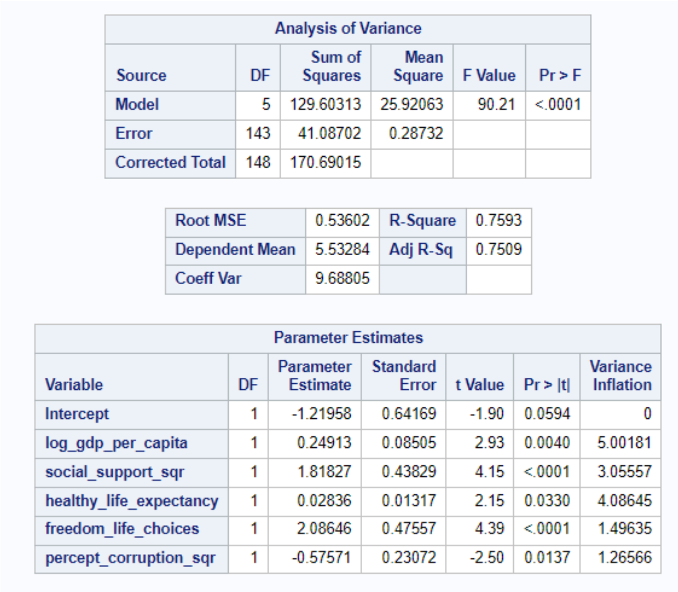
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	130.35666	21.72611	76.49	<.0001
Error	142	40.33349	0.28404		
Corrected Total	148	170.69015			

Root MSE	0.53295	R-Square	0.7637
Dependent Mean	5.53284	Adj R-Sq	0.7537
Coeff Var	9.63254		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-4.45341	2.08543	-2.14	0.0344	0
log_gdp_per_capita	1	0.24732	0.08457	2.92	0.0040	5.00267
social_support	1	-0.81957	0.50318	-1.63	0.1056	131.92642
social_support_sqr	1	6.48613	2.89882	2.24	0.0268	135.21077
healthy_life_expectancy	1	0.02809	0.01310	2.14	0.0337	4.08710
freedom_life_choices	1	2.14819	0.47437	4.53	<.0001	1.50597
percept_corruption_sqr	1	-0.46934	0.23851	-1.97	0.0510	1.36824

Since it didn’t help with VIF’s I decided to drop social support regressor and go with social support squared. I also tried the BoxCox transformation, but it didn’t help, so I decided not to change it. So, the model is:

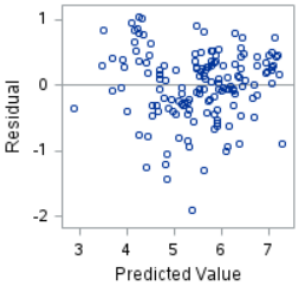


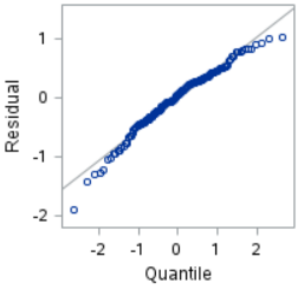
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	129.60313	25.92063	90.21	<.0001
Error	143	41.08702	0.28732		
Corrected Total	148	170.69015			

Root MSE	0.53602	R-Square	0.7593
Dependent Mean	5.53284	Adj R-Sq	0.7509
Coeff Var	9.68805		

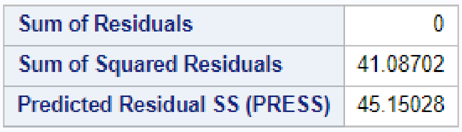
Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-1.21958	0.64169	-1.90	0.0594	0
log_gdp_per_capita	1	0.24913	0.08505	2.93	0.0040	5.00181
social_support_sqr	1	1.81827	0.43829	4.15	<.0001	3.05557
healthy_life_expectancy	1	0.02836	0.01317	2.15	0.0330	4.08645
freedom_life_choices	1	2.08646	0.47557	4.39	<.0001	1.49635
percept_corruption_sqr	1	-0.57571	0.23072	-2.50	0.0137	1.26566

We see that VIFs for gdp and healthy life expectancy is still high if compared to others. All regressors are significant, R^2 and R^2_{Adj} improved a little. MSE improved a little.



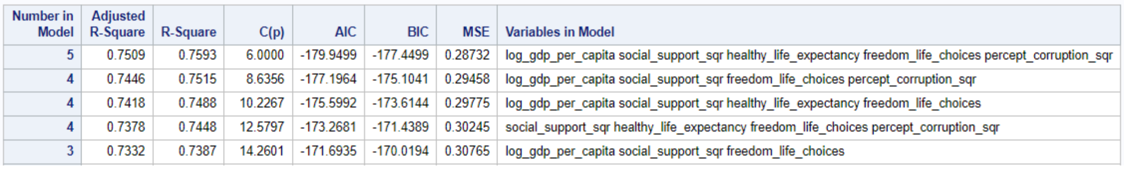


Constant variance assumption is still not satisfied, as well as linearity assumption. NPP plot still shows some skewness.



Sum of Residuals	0
Sum of Squared Residuals	41.08702
Predicted Residual SS (PRESS)	45.15028

PRESS statistics improved a little, as well as R^2_{Pred} = 73.55%.



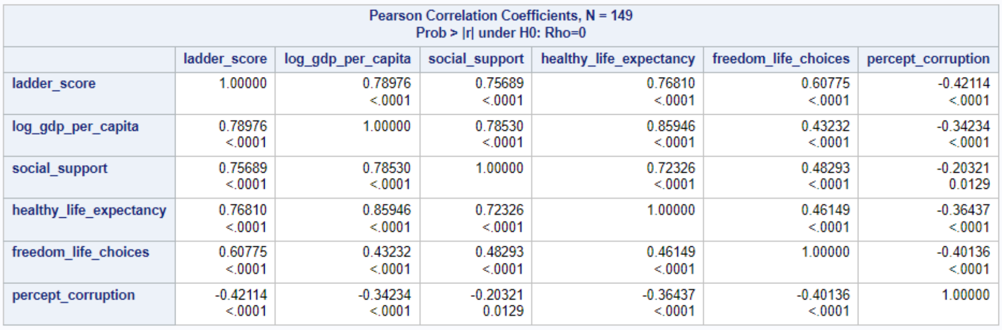
Number in Model	Adjusted R-Square	R-Square	C(p)	AIC	BIC	MSE	Variables in Model
5	0.7509	0.7593	6.0000	-179.9499	-177.4499	0.28732	log_gdp_per_capita social_support_sqr healthy_life_expectancy freedom_life_choices percept_corruption_sqr
4	0.7446	0.7515	8.6356	-177.1964	-175.1041	0.29458	log_gdp_per_capita social_support_sqr freedom_life_choices percept_corruption_sqr
4	0.7418	0.7488	10.2267	-175.5992	-173.6144	0.29775	log_gdp_per_capita social_support_sqr healthy_life_expectancy freedom_life_choices
4	0.7378	0.7448	12.5797	-173.2681	-171.4389	0.30245	social_support_sqr healthy_life_expectancy freedom_life_choices percept_corruption_sqr
3	0.7332	0.7387	14.2601	-171.6935	-170.0194	0.30765	log_gdp_per_capita social_support_sqr freedom_life_choices

C_p statistics shows that our model is unbiased now and it’s lower since we have 1 less regressor, and we see some small improvements in AIC and BIC statistics.

Since I didn’t get any sufficient improvements, I can go with the original data. I also tried to normalize gdp and healthy life expectancy regressors, but it didn’t change anything.

Data transformation through adding cross regressors

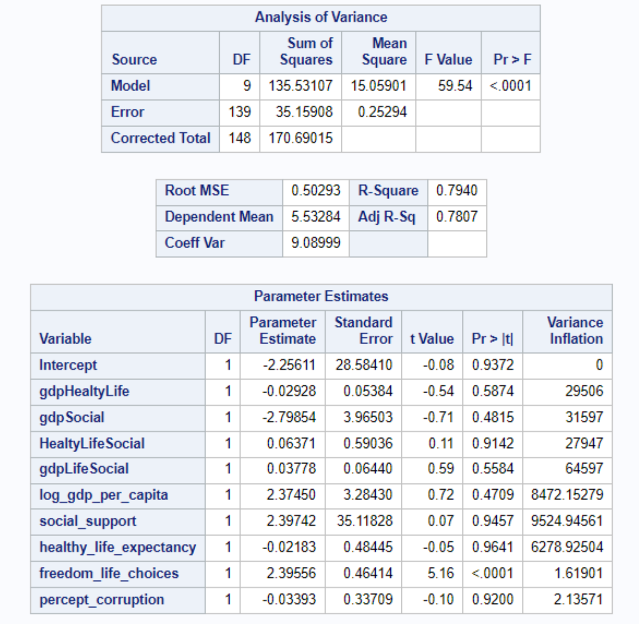
Then, I tried to check the correlation matrix.



Pearson Correlation Coefficients, N = 149
Prob > |r| under H0: Rho=0

	ladder_score	log_gdp_per_capita	social_support	healthy_life_expectancy	freedom_life_choices	percept_corruption
ladder_score	1.00000	0.78976 <.0001	0.75689 <.0001	0.76810 <.0001	0.60775 <.0001	-0.42114 <.0001
log_gdp_per_capita	0.78976 <.0001	1.00000	0.78530 <.0001	0.85946 <.0001	0.43232 <.0001	-0.34234 <.0001
social_support	0.75689 <.0001	0.78530 <.0001	1.00000	0.72326 <.0001	0.48293 <.0001	-0.20321 0.0129
healthy_life_expectancy	0.76810 <.0001	0.85946 <.0001	0.72326 <.0001	1.00000	0.46149 <.0001	-0.36437 <.0001
freedom_life_choices	0.60775 <.0001	0.43232 <.0001	0.48293 <.0001	0.46149 <.0001	1.00000	-0.40136 <.0001
percept_corruption	-0.42114 <.0001	-0.34234 <.0001	-0.20321 0.0129	-0.36437 <.0001	-0.40136 <.0001	1.00000

I can see that there is a huge correlation between various variables. Especially, gdp has huge correlation with social support and healthy life expectancy. All three of them are highly correlated. I decided to add products of gdp and social support, gdp and healthy life expectancy, healthy life expectancy and social support, all three of them. So, I got the model:



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	135.53107	15.05901	59.54	<.0001
Error	139	35.15908	0.25294		
Corrected Total	148	170.69015			

Root MSE	0.50293	R-Square	0.7940
Dependent Mean	5.53284	Adj R-Sq	0.7807
Coeff Var	9.08999		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-2.25611	28.58410	-0.08	0.9372	0
gdpHealtyLife	1	-0.02928	0.05384	-0.54	0.5874	29506
gdpSocial	1	-2.79854	3.96503	-0.71	0.4815	31597
HealtyLifeSocial	1	0.06371	0.59036	0.11	0.9142	27947
gdpLifeSocial	1	0.03778	0.06440	0.59	0.5584	64597
log_gdp_per_capita	1	2.37450	3.28430	0.72	0.4709	8472.15279
social_support	1	2.39742	35.11828	0.07	0.9457	9524.94561
healthy_life_expectancy	1	-0.02183	0.48445	-0.05	0.9641	6278.92504
freedom_life_choices	1	2.39556	0.46414	5.16	<.0001	1.61901
percept_corruption	1	-0.03393	0.33709	-0.10	0.9200	2.13571

Adjusted R^2 improved, but we have huge VIFs, so we have to choose the regressors, that will explain the model and will not have huge VIFs. I check this using extra sum of squares method.

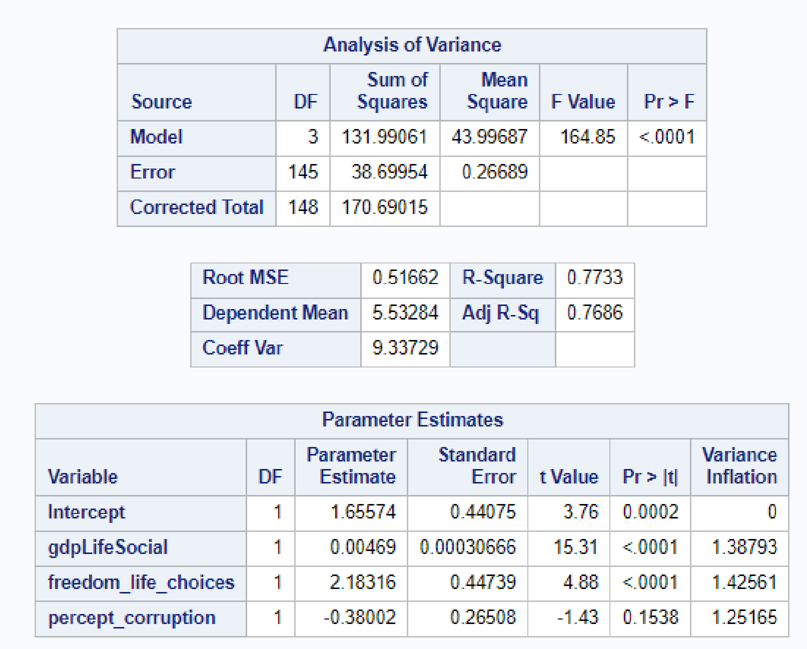
Since the regressor that combines all three of the regressors, I decided to check whether other products and gdp, healthy life expectancy and social support are significant.

H_0: gdpHealtyLife = gdpSocial = HealtyLifeSocial = log_gdp_per_capita = social_support = healthy_life_expectancy = 0.

H_1: At least one of them is not equal to 0.

$$F_0 = \frac{(35.15908 - 38.69954)/6}{0.25294} = -2.3328$$

Since $F_{0.025,\, 6,10}$ = 4.07, we can't reject our hypothesis H_0. So, let's try the model without them.

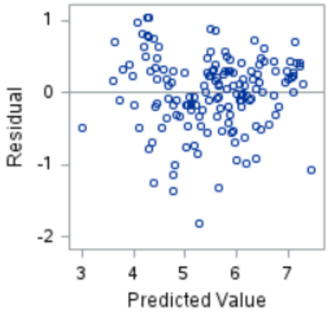


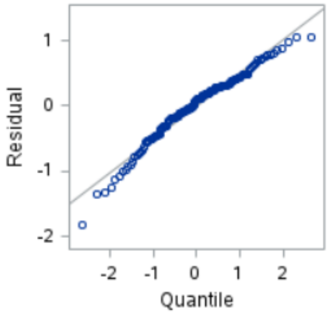
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	131.99061	43.99687	164.85	<.0001
Error	145	38.69954	0.26689		
Corrected Total	148	170.69015			

Root MSE	0.51662	R-Square	0.7733
Dependent Mean	5.53284	Adj R-Sq	0.7686
Coeff Var	9.33729		

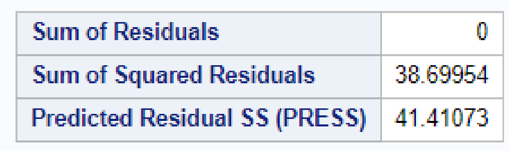
Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	1.65574	0.44075	3.76	0.0002	0
gdpLifeSocial	1	0.00469	0.00030666	15.31	<.0001	1.38793
freedom_life_choices	1	2.18316	0.44739	4.88	<.0001	1.42561
percept_corruption	1	-0.38002	0.26508	-1.43	0.1538	1.25165

We see that R^2 and R^2_{Adj} improved, MSE also improved and we don't have huge VIFs.



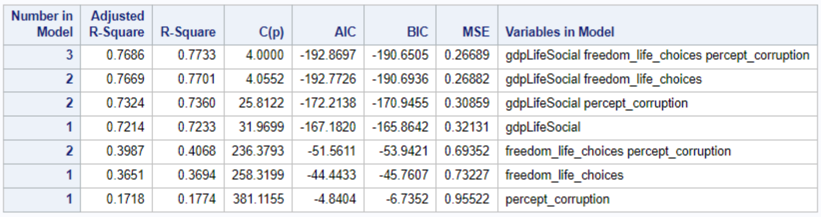


Constant variance assumption improved as well as we don’t see huge nonlinear pattern anymore. NPP improved.



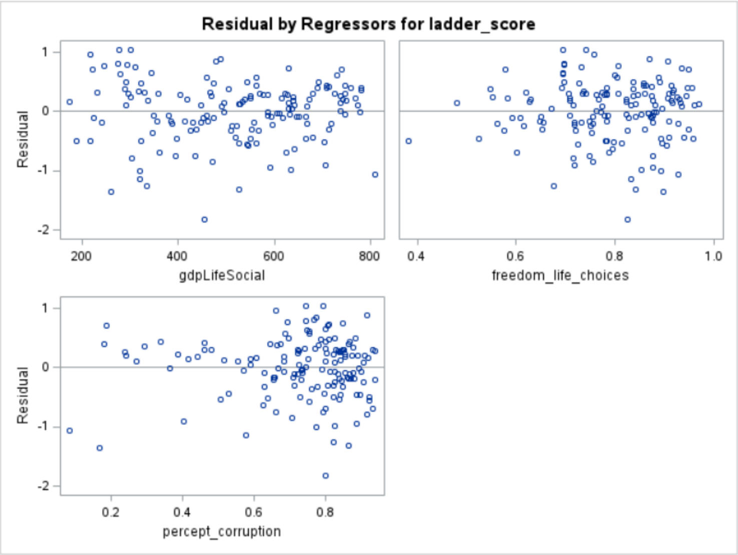
Sum of Residuals	0
Sum of Squared Residuals	38.69954
Predicted Residual SS (PRESS)	41.41073

PRESS statistics became lower, and now our R^2_{Pred} = 75.73% which is by couple percent better than initial model.

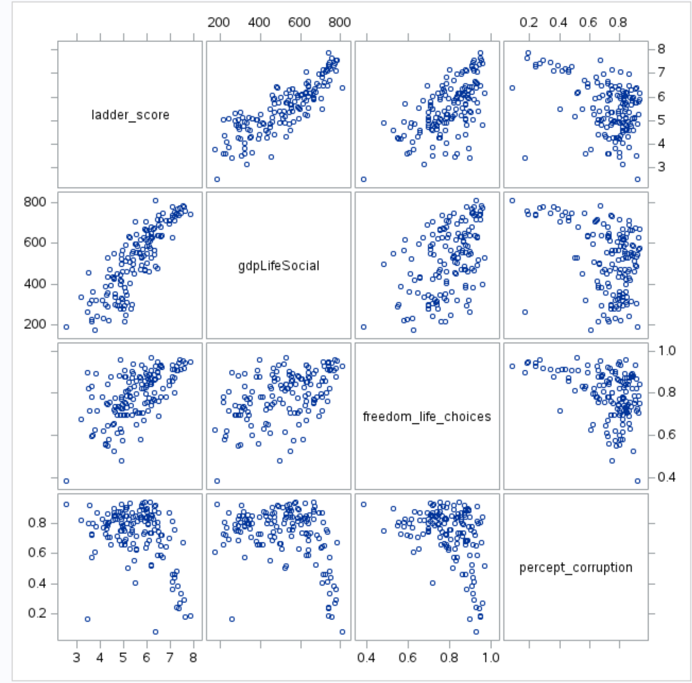


Number in Model	Adjusted R-Square	R-Square	C(p)	AIC	BIC	MSE	Variables in Model
3	0.7686	0.7733	4.0000	-192.8697	-190.6505	0.26689	gdpLifeSocial freedom_life_choices percept_corruption
2	0.7669	0.7701	4.0552	-192.7726	-190.6936	0.26882	gdpLifeSocial freedom_life_choices
2	0.7324	0.7360	25.8122	-172.2138	-170.9455	0.30859	gdpLifeSocial percept_corruption
1	0.7214	0.7233	31.9699	-167.1820	-165.8642	0.32131	gdpLifeSocial
2	0.3987	0.4068	236.3793	-51.5611	-53.9421	0.69352	freedom_life_choices percept_corruption
1	0.3651	0.3694	258.3199	-44.4433	-45.7607	0.73227	freedom_life_choices
1	0.1718	0.1774	381.1155	-4.8404	-6.7352	0.95522	percept_corruption

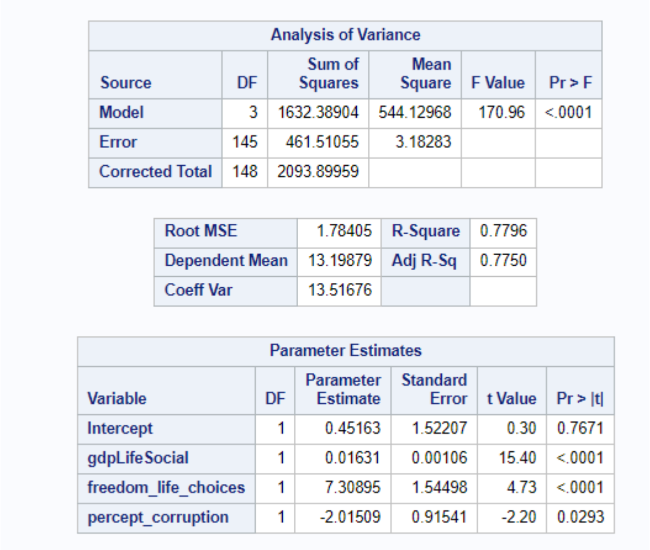
All the statistics became better, the model is unbiased. MSE is smaller, as well as AIC and BIC, while R^2 got better.



We still see some problems with both freedom of life choices and perception of corruption. Let’s look at the scatter plot.



We see that there is an obvious nonlinear pattern. Let’s try to use BoxCox method and see what will happen.

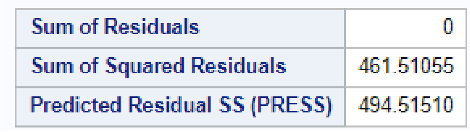


Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1632.38904	544.12968	170.96	<.0001
Error	145	461.51055	3.18283		
Corrected Total	148	2093.89959			

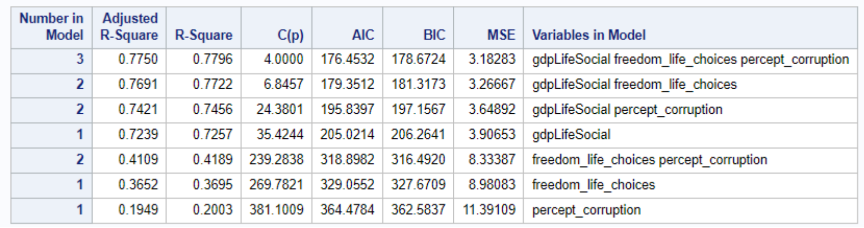
Root MSE	1.78405	R-Square	0.7796
Dependent Mean	13.19879	Adj R-Sq	0.7750
Coeff Var	13.51676		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.45163	1.52207	0.30	0.7671
gdpLifeSocial	1	0.01631	0.00106	15.40	<.0001
freedom_life_choices	1	7.30895	1.54498	4.73	<.0001
percept_corruption	1	-2.01509	0.91541	-2.20	0.0293

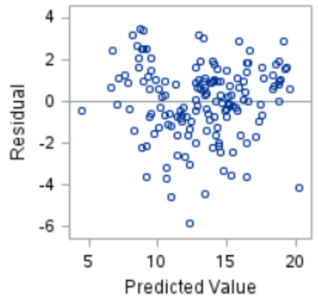
We see that though R^2 and R^2_{Adj} became better, intercept became insignificant, Sum of Squares increased, which means PRESS will increase as well as AIC and BIC.

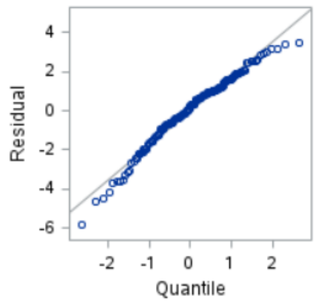


Sum of Residuals	0
Sum of Squared Residuals	461.51055
Predicted Residual SS (PRESS)	494.51510



Number in Model	Adjusted R-Square	R-Square	C(p)	AIC	BIC	MSE	Variables in Model
3	0.7750	0.7796	4.0000	176.4532	178.6724	3.18283	gdpLifeSocial freedom_life_choices percept_corruption
2	0.7691	0.7722	6.8457	179.3512	181.3173	3.26667	gdpLifeSocial freedom_life_choices
2	0.7421	0.7456	24.3801	195.8397	197.1567	3.64892	gdpLifeSocial percept_corruption
1	0.7239	0.7257	35.4244	205.0214	206.2641	3.90653	gdpLifeSocial
2	0.4109	0.4189	239.2838	318.8982	316.4920	8.33387	freedom_life_choices percept_corruption
1	0.3652	0.3695	269.7821	329.0552	327.6709	8.98083	freedom_life_choices
1	0.1949	0.2003	381.1009	364.4784	362.5837	11.39109	percept_corruption

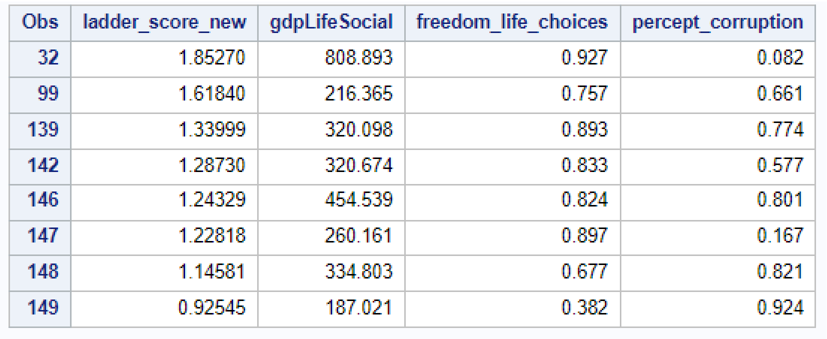




Didn't help with constant variance and NPP. So, since it didn't help us, we will not use BoxCox. I tried to do different types of transformation of corruption and freedom of life choices but didn't succeed.

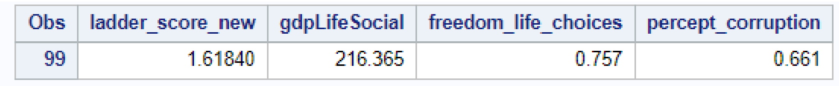
Detecting outliers

Cook’s D outliers:



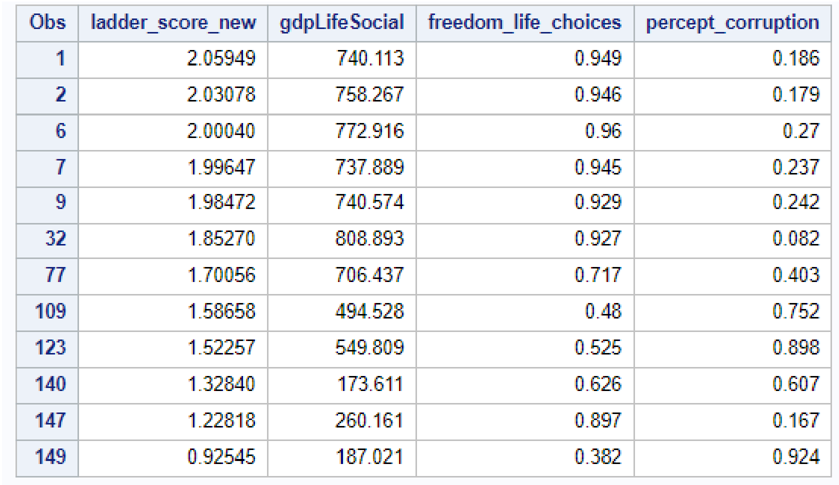
Obs	ladder_score_new	gdpLifeSocial	freedom_life_choices	percept_corruption
32	1.85270	808.893	0.927	0.082
99	1.61840	216.365	0.757	0.661
139	1.33999	320.098	0.893	0.774
142	1.28730	320.674	0.833	0.577
146	1.24329	454.539	0.824	0.801
147	1.22818	260.161	0.897	0.167
148	1.14581	334.803	0.677	0.821
149	0.92545	187.021	0.382	0.924

DFFIT outliers:



Obs	ladder_score_new	gdpLifeSocial	freedom_life_choices	percept_corruption
99	1.61840	216.365	0.757	0.661

Hat diagonal influential:



Obs	ladder_score_new	gdpLifeSocial	freedom_life_choices	percept_corruption
1	2.05949	740.113	0.949	0.186
2	2.03078	758.267	0.946	0.179
6	2.00040	772.916	0.96	0.27
7	1.99647	737.889	0.945	0.237
9	1.98472	740.574	0.929	0.242
32	1.85270	808.893	0.927	0.082
77	1.70056	706.437	0.717	0.403
109	1.58658	494.528	0.48	0.752
123	1.52257	549.809	0.525	0.898
140	1.32840	173.611	0.626	0.607
147	1.22818	260.161	0.897	0.167
149	0.92545	187.021	0.382	0.924

Appendix (SAS code)

```
libname final base "/home/u63145576/Final project";

Options validvarname=v7;

proc import datafile='/home/u63145576/Final project/world-happiness-report-2021.csv'

                              DBMS=csv

                              out=final.happiness

                              replace;

                              getnames=yes;

run;


data final.happiness(rename=('Ladder score'n= ladder_score 'Standard error of ladder
score'n=std_ladder

                                                            'Logged GDP per capita'n = log_gdp_per_capita
'Social support'n = social_support

                                                            'Healthy life expectancy'n =
healthy_life_expectancy

                                                            'Freedom to make life choices'n =
freedom_life_choices

                                                            'Perceptions of corruption'n =
percept_corruption));/*rename sas converted names to some convenient names*/

        set final.happiness;

run;

proc contents data=final.happiness;

run;

proc print data=final.happiness;

run;


*--------------------initial model----------------------------;

proc reg data=final.happiness;

        model ladder_score = log_gdp_per_capita

                                                            social_support  healthy_life_expectancy
        freedom_life_choices

                                                            Generosity      percept_corruption/influence
press;

run;
```

```
proc reg data=final.happiness;
        model ladder_score = log_gdp_per_capita
                                                        social_support  healthy_life_expectancy
        freedom_life_choices
                                                        Generosity      percept_corruption/vif ;
run;
proc reg data=final.happiness;
        model ladder_score = log_gdp_per_capita
                                                        social_support  healthy_life_expectancy
        freedom_life_choices
                                                        Generosity
        percept_corruption/selection=adjrsq mse aic bic cp;
run;
*----------------Initial model--------------------------;
*-------------Backward elimination---------------------------------;
proc reg data=final.happiness;
        model ladder_score = log_gdp_per_capita
                                                        social_support  healthy_life_expectancy
        freedom_life_choices
                                                        Generosity      percept_corruption
/selection=backward;
run;


*model after backward elimination;
proc reg data=final.happiness;
        model ladder_score = log_gdp_per_capita
                                                        social_support  healthy_life_expectancy
        freedom_life_choices
                                                        Generosity percept_corruption /influence press;
run;


*--------------Forward selection---------------;
proc reg data=final.happiness;
        model ladder_score = log_gdp_per_capita
```

```
                                                  social_support  healthy_life_expectancy
       freedom_life_choices
                                                  Generosity      percept_corruption
/selection=forward;
run;
proc reg data=final.happiness;
       model ladder_score = log_gdp_per_capita
                                                  social_support  healthy_life_expectancy
       freedom_life_choices
                                                  percept_corruption /influence press;
run;

*---------------------Stepwise----------------------------;
*full model;
proc reg data=final.happiness;
       model ladder_score = log_gdp_per_capita
                                                  social_support  healthy_life_expectancy
       freedom_life_choices
                                                  Generosity      percept_corruption
/selection=stepwise;
run;

proc reg data=final.happiness;
       model ladder_score = log_gdp_per_capita
                                                  social_support  healthy_life_expectancy
       freedom_life_choices
                                                  percept_corruption /vif;
run;

*---------------------Ten 10 best models----------------------------------------------------------;
proc reg data=final.happiness;
       model ladder_score = log_gdp_per_capita
                                                  social_support  healthy_life_expectancy
       freedom_life_choices
```

```
                                                    Generosity      percept_corruption
/selection=cp best=10;

run;

proc reg data=final.happiness;

        model ladder_score = log_gdp_per_capita

                                                    social_support  healthy_life_expectancy
        freedom_life_choices

                                                    Generosity      percept_corruption
/selection=adjrsq mse aic bic;

run;

proc reg data=final.happiness;

        model ladder_score =  log_gdp_per_capita

                                                    social_support  healthy_life_expectancy
        freedom_life_choices

                                                    percept_corruption /influence press;

run;


*-----------------working with the chosen regressors---------------------------------;

proc reg data=final.happiness;

        model ladder_score = log_gdp_per_capita

                                                    social_support  healthy_life_expectancy
        freedom_life_choices

                                                    percept_corruption/vif;

run;


*-------------Check for constant variance------------------------------;

proc reg data=final.happiness plots = residualbypredicted;

   ods select residualbypredicted;

   model ladder_score = log_gdp_per_capita social_support healthy_life_expectancy

                                             freedom_life_choices percept_corruption;

run;

proc reg data=final.happiness plots = residualbypredicted;

   ods select residualbypredicted;

   model ladder_score = log_gdp_per_capita;
```

```
run;
proc reg data=final.happiness plots = residualbypredicted;
    ods select residualbypredicted;
    model ladder_score = social_support;
run;
proc reg data=final.happiness plots = residualbypredicted;
    ods select residualbypredicted;
    model ladder_score = healthy_life_expectancy;
run;
proc reg data=final.happiness plots = residualbypredicted;
    ods select residualbypredicted;
    model ladder_score = freedom_life_choices;
run;
proc reg data=final.happiness plots = residualbypredicted;
    ods select residualbypredicted;
    model ladder_score = percept_corruption;
run;
*---------------Transformations----------------------;
data final.happiness3;
            set final.happiness;
            social_support_sqr = social_support**2;
            percept_corruption_sqr = (percept_corruption)**2;
            ladder_score_new = (ladder_score)**(1.5);
run;
*---------res vs pred after transformations----------;
proc reg data=final.happiness3 plots = residualbypredicted;
    ods select residualbypredicted;
    model ladder_score = percept_corruption_sqr;
run;
proc reg data=final.happiness3 plots = residualbypredicted;
    ods select residualbypredicted;
    model ladder_score = social_support social_support_sqr;
```

```
run;

proc reg data=final.happiness3 plots = residualbypredicted;

    ods select residualbypredicted;

    model ladder_score = log_gdp_per_capita social_support_sqr healthy_life_expectancy

                                                                freedom_life_choices percept_corruption;

run;

*to fix social we need to add social^2--------------------;

proc reg data=final.happiness3 plots = residualbypredicted;

    ods select residualbypredicted;

    model ladder_score =  social_support social_support_sqr;

run;


proc reg data=final.happiness3;

            model ladder_score_new = log_gdp_per_capita

                                                                        social_support social_support_sqr
healthy_life_expectancy

                                                                        freedom_life_choices percept_corruption_sqr
/vif;

run;

*----------trying to normalize social----------------;

proc means data=final.happiness3 Mean StdDev ndec=3;

  var social_support;

run;

proc stdize data=final.happiness3 out=normalized_data;

  var social_support percept_corruption;

run;

proc means data=normalized_data Mean StdDev ndec=3;

  var social_support;

run;

proc reg data=normalized_data plots = residualbypredicted;

    ods select residualbypredicted;

    model ladder_score =   social_support_sqr;

run;
```

```
proc reg data=normalized_data plots = residualbypredicted;
   ods select residualbypredicted;
   model ladder_score = percept_corruption_sqr;
run;
proc reg data=normalized_data;
        model ladder_score = log_gdp_per_capita
                                                        social_support social_support_sqr
healthy_life_expectancy
                                                        freedom_life_choices percept_corruption_sqr
/vif;
run;
proc reg data=normalized_data;
        model ladder_score_new = log_gdp_per_capita
                                                         social_support_sqr healthy_life_expectancy
                                                        freedom_life_choices percept_corruption_sqr
/vif;
run;


proc reg data=final.happiness3;
        model ladder_score_new = log_gdp_per_capita
                                                         social_support_sqr healthy_life_expectancy
                                                        freedom_life_choices percept_corruption_sqr
/vif;
run;
proc reg data=final.happiness3;
        model ladder_score_new = log_gdp_per_capita
                                                         social_support_sqr healthy_life_expectancy
                                                        freedom_life_choices percept_corruption_sqr
/selection=adjrsq mse aic bic cp;
run;
proc reg data=final.happiness3;
        model ladder_score_new = log_gdp_per_capita
                                                         social_support_sqr healthy_life_expectancy
```

```
                                                freedom_life_choices percept_corruption_sqr
/selection=adjrsq mse aic bic;
run;
proc transreg data=final.happiness3 test;
      model BoxCox(ladder_score) = identity(log_gdp_per_capita
                                                 social_support_sqr healthy_life_expectancy
                                                freedom_life_choices percept_corruption_sqr);
run;
*-----------deleting social because of vif----------------------;
proc reg data=final.happiness3;
      model ladder_score = log_gdp_per_capita
                                                 social_support_sqr healthy_life_expectancy
                                                freedom_life_choices percept_corruption_sqr
/vif;
run;
proc reg data=final.happiness3;
      model ladder_score = log_gdp_per_capita
                                                 social_support_sqr healthy_life_expectancy
                                                freedom_life_choices percept_corruption_sqr
/selection=adjrsq mse aic bic cp;
run;
proc reg data=final.happiness3;
      model ladder_score = log_gdp_per_capita
                                                 social_support_sqr healthy_life_expectancy
                                                freedom_life_choices percept_corruption_sqr
/selection=cp best=10;
run;
proc reg data=final.happiness3;
      model ladder_score = log_gdp_per_capita
                                                 social_support_sqr healthy_life_expectancy
                                                freedom_life_choices percept_corruption_sqr
/influence press;
run;
```

```
proc transreg data=final.happiness3 test;
            model BoxCox(ladder_score) = identity(log_gdp_per_capita
                                                                          social_support_sqr healthy_life_expectancy
                                                                          freedom_life_choices percept_corruption_sqr);
run;
*----------press increased rapidly and didn't help, so leave response the same----;
proc reg data=final.happiness3;
            model ladder_score_new = log_gdp_per_capita
                                                                          social_support_sqr healthy_life_expectancy
                                                                          freedom_life_choices percept_corruption_sqr
/influence press;
run;
proc reg data=final.happiness3;
            model ladder_score_new = log_gdp_per_capita
                                                                          social_support_sqr healthy_life_expectancy
                                                                          freedom_life_choices percept_corruption_sqr
/vif;
run;
proc reg data=final.happiness3;
            model ladder_score_new = log_gdp_per_capita
                                                                          social_support_sqr healthy_life_expectancy
                                                                          freedom_life_choices percept_corruption_sqr
/selection=adjrsq mse aic bic;
run;
proc reg data=final.happiness3;
            model ladder_score_new = log_gdp_per_capita
                                                                          social_support_sqr healthy_life_expectancy
                                                                          freedom_life_choices percept_corruption_sqr
/selection=cp best=10;
run;
proc reg data=final.happiness3;
```

```
	model ladder_score_new = log_gdp_per_capita
								     social_support_sqr healthy_life_expectancy
								    freedom_life_choices percept_corruption_sqr
/influence press;
run;




*-------------the best one without boxcox------------------------------;
*---------------------Final model----------------------;
proc reg data=final.happiness3;
	model ladder_score = log_gdp_per_capita
								     social_support_sqr healthy_life_expectancy
								    freedom_life_choices percept_corruption_sqr
/vif;
run;
proc reg data=final.happiness3;
	model ladder_score = log_gdp_per_capita
								     social_support_sqr healthy_life_expectancy
								    freedom_life_choices percept_corruption_sqr
/selection=adjrsq mse aic bic cp;
run;
proc reg data=final.happiness3;
	model ladder_score = log_gdp_per_capita
								     social_support_sqr healthy_life_expectancy
								    freedom_life_choices percept_corruption_sqr
/influence press;
run;
proc reg data=final.happiness3;
	model ladder_score = log_gdp_per_capita
								     social_support_sqr healthy_life_expectancy
								    freedom_life_choices percept_corruption_sqr /r;

run;
```

```
*---------tried to normalize---------------------------;


proc stdize data=final.happiness3 out=normalized_data;
  var log_gdp_per_capita healthy_life_expectancy;
run;


proc reg data=normalized_data;
          model ladder_score = log_gdp_per_capita
                                                            social_support healthy_life_expectancy
                                                           freedom_life_choices percept_corruption /vif;
run;


*------------Looking at the scatter and correlation matrix-----------;
proc sgscatter data=final.happiness;
           matrix ladder_score log_gdp_per_capita
                                                           social_support  healthy_life_expectancy
          freedom_life_choices
                                                                     percept_corruption generosity;
run;


proc sgscatter data=final.happiness2;
           matrix ladder_score log_gdp_per_capita
                                                           social_support  healthy_life_expectancy
          freedom_life_choices
                                                                     percept_corruption_sqr;
run;
proc corr data=final.happiness;
          var ladder_score log_gdp_per_capita
                                                           social_support  healthy_life_expectancy
          freedom_life_choices
```

```
                                                            percept_corruption;
run;
data final.happinessAddingCross;
        set final.happiness;
        gdpHealtyLife = log_gdp_per_capita * healthy_life_expectancy;
        gdpSocial = log_gdp_per_capita*social_support;
        HealtyLifeSocial = healthy_life_expectancy*social_support;
        gdpLifeSocial = log_gdp_per_capita*healthy_life_expectancy*social_support;
run;
proc corr data=final.happinessAddingCross;
        var ladder_score  gdpHealtyLife gdpSocial HealtyLifeSocial gdpLifeSocial
                                                log_gdp_per_capita      social_support
        healthy_life_expectancy         freedom_life_choices
                                                            percept_corruption;
run;
proc reg data=final.happinessAddingCross;
        model ladder_score =  gdpHealtyLife gdpSocial HealtyLifeSocial gdpLifeSocial
                                                log_gdp_per_capita      social_support
        healthy_life_expectancy         freedom_life_choices
                                                            percept_corruption/vif;
run;


proc reg data=final.happinessAddingCross;
        model ladder_score =     gdpLifeSocial
                                                                  freedom_life_choices
                                                            percept_corruption/vif;
run;
proc reg data=final.happinessAddingCross;
        model ladder_score =     gdpLifeSocial
                                                                  freedom_life_choices
                                                            percept_corruption/selection=adjrsq
mse aic bic cp;
run;
```

```
proc reg data=final.happinessAddingCross;
        model ladder_score =    gdpLifeSocial
                                                                  freedom_life_choices
                                                          percept_corruption/influence press;
run;

*------------problems with corruption-------------------;
proc sgscatter data=final.happinessAddingCross;
         matrix ladder_score gdpLifeSocial
                                                          freedom_life_choices
                                                          percept_corruption;
run;
proc stdize data=final.happinessAddingCross out=normalized_data;
  var  percept_corruption;
run;
proc sgscatter data=normalized_data;
         matrix ladder_score gdpLifeSocial
                                                                  freedom_life_choices
                                                          percept_corruption;
run;
proc transreg data=final.happinessaddingcross test;
        model BoxCox(ladder_score)=identity(gdpLifeSocial
                                                  freedom_life_choices
                                                  percept_corruption);
run;
data final.happinessTransfAfterCross;
        set final.happiness;
        gdpHealtyLife = log_gdp_per_capita * healthy_life_expectancy;
        gdpSocial = log_gdp_per_capita*social_support;
        HealtyLifeSocial = healthy_life_expectancy*social_support;
        gdpLifeSocial = log_gdp_per_capita*healthy_life_expectancy*social_support;
        percept_corruption_new = (percept_corruption);
```

```
            freedom_life_choices_new = log(freedom_life_choices);
            ladder_score_new = log(ladder_score);
run;
proc reg data=final.happinessTransfAfterCross;
            model ladder_score_new = gdpLifeSocial
                                                                        freedom_life_choices
                                                                        percept_corruption;
run;
proc reg data=final.happinessTransfAfterCross;
            model ladder_score_new = gdpLifeSocial
                                                                        freedom_life_choices
                                                                        percept_corruption/influence press;
run;
proc reg data=final.happinessTransfAfterCross;
            model ladder_score_new = gdpLifeSocial
                                                                        freedom_life_choices
                                                                        percept_corruption/selection=adjrsq mse aic bic
cp;
run;
proc sgscatter data=final.happinessTransfAfterCross;
             matrix ladder_score_new gdpLifeSocial
                                                                                        freedom_life_choices_new
                                                                                percept_corruption_new;
run;

proc reg data=final.happinessTransfAfterCross;
            model ladder_score_new = gdpLifeSocial
                                                                        freedom_life_choices
                                                                        percept_corruption;
run;

*---------outliers------------------;
```

```
proc reg data=final.happinessTransfAfterCross;
model ladder_score_new = gdpLifeSocial
                                                        freedom_life_choices
                                                        percept_corruption / stb
clb;
output out=stdres p= predict r = resid rstudent=r h=lev
cookd=cookd dffits=dffit;
run;
*---------CooksD outliers-------------------;
proc print data=stdres;
        where cookd>(4/149);
        var ladder_score_new gdpLifeSocial
                                                        freedom_life_choices
                                                        percept_corruption;
run;
*----------DFFIT------------------;
proc print data=stdres;
        where dffit> abs(2*((4/149)**0.5));
        var ladder_score_new gdpLifeSocial
                                                        freedom_life_choices
                                                        percept_corruption;
run;
*------------H hat------------------;
proc print data=stdres;
        where lev> 2*4/149;
        var ladder_score_new gdpLifeSocial
                                                        freedom_life_choices
                                                        percept_corruption;
run;
```

```
*---------trying code to do weighted least squares----------------;
/* Weighted Least Squares as an Adjustment */
proc reg data=final.happinessTransfAfterCross;
 model ladder_score = gdpLifeSocial
                                                       freedom_life_choices
                                                       percept_corruption;
 output out=WORK.PRED r=residual;
run;

data work.resid;
 set work.pred;
 absresid=abs(residual);
 sqresid=residual**2;



proc reg data=work.resid;
 model ladder_score = gdpLifeSocial
                                                       freedom_life_choices
                                                       percept_corruption;
 output out=WORK.s_weights p=s_hat;
 model ladder_score = gdpLifeSocial
                                                       freedom_life_choices
                                                       percept_corruption;
 output out=WORK.v_weights p=v_hat;
run;
** compute the weights using the estimated standard deviations**;
data work.s_weights;
set work.s_weights;
s_weight=1/(s_hat**2);
label s_weight = "weights using absolute residuals";
** compute the weights using the estimated variances**;
data work.v_weights;
```

```
set work.v_weights;
v_weight=1/v_hat;
label v_weight = "weights using squared residuals";
** Do the weighted least squares using the weights from the estimated
standard deviation**;
proc reg data=work.s_weights;
weight s_weight;
model ladder_score = gdpLifeSocial
                                                      freedom_life_choices
                                                      percept_corruption;
run;
proc reg data=work.v_weights;
weight v_weight;
model ladder_score = gdpLifeSocial
                                                      freedom_life_choices
                                                      percept_corruption;
run;
```

References

1. Schreiber-Gregory, D. (n.d.). *Logistic and linear regression assumptions: Violation recognition and ...* Retrieved May 2, 2023, from https://www.lexjansen.com/mwsug/2018/AA/MWSUG-2018-AA-91.pdf
2. Singh, A. (2021, March 22). *World happiness report 2021*. Kaggle. Retrieved May 1, 2023, from https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021
3. Helliwell, J. F., Layard, R., Sachs, J. D., Neve, J.-E. D., Aknin, L. B., & Wang, S. (2023, March 20). *World happiness report 2023*. The World Happiness Report. Retrieved May 1, 2023, from https://worldhappiness.report/ed/2023/#appendices-and-data
4. *John F. Helliwell, Richard Layard, Jeffrey D. Sachs, jan-emmanuel de ...* (n.d.). Retrieved May 2, 2023, from https://happiness-report.s3.amazonaws.com/2022/WHR+22.pdf