Report on Statistical Analysis using Multiple Linear Regression

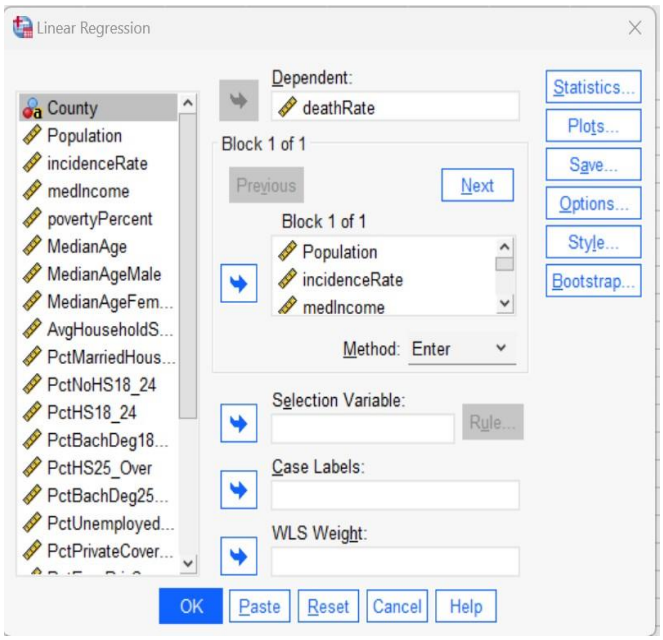
Statistics Assignment CA

Amal Sunny
Master of Science in Data Analytics
National College of Ireland
Dublin, Ireland
x22169806@student.ncirl.ie

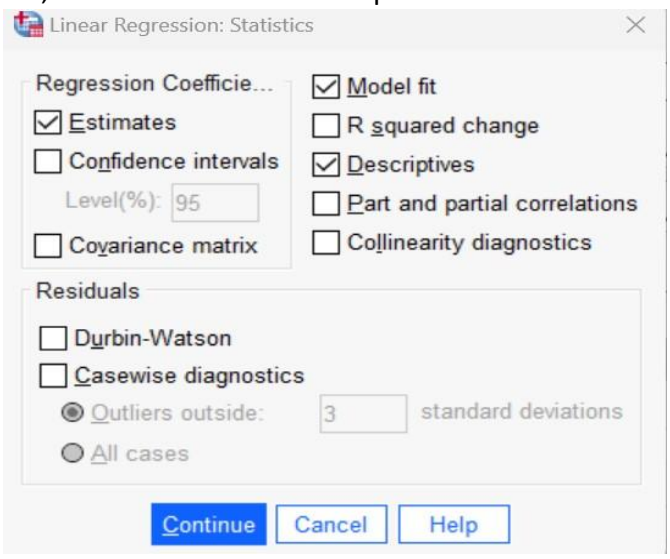
Statistics CA using SPSS.

The Statistical Package for the Social Sciences (SPSS) is a software tool commonly utilized in social science research, particularly in the domains of psychology, sociology, economics, and related disciplines. SPSS facilitates the sorting, analysis, and comprehension of extensive data sets using a diverse range of statistical methods. These methods comprise descriptive statistics such as frequency distributions, means, and standard deviations, as well as inferential statistics like regression analysis, analysis of variance (ANOVA), and chi-square tests.

- To begin, the initial action is to import the dataset named cancer.csv. The response variable within this dataset is the Death Rate, while all of the socio-economic data that is available are considered as the independent variables.
- Once the data has been imported, we can obtain an overview of it by utilizing descriptive statistics. This overview will encompass the mean, median, standard deviation, and range of all variables within the dataset.
- Assigning the dependent and independent variables:



- To construct a regression model, we need to select Descriptives to obtain an overview of the data, and then Model fit and estimates to assess the model's fit.



- Descriptive Statistics: Outline of the data using descriptive statistics, including mean, standard deviation, and range.

Descriptive Statistics

	Mean	Std. Deviation	N
deathRate	178.664	27.7515	3047
Population	102637.37	329059.221	3047
incidenceRate	445.654	57.4566	3047
medIncome	47063.28	12040.091	3047
povertyPercent	16.878	6.4091	3047
MedianAge	45.272	45.3045	3047
MedianAgeMale	39.571	5.2260	3047
MedianAgeFemale	42.145	5.2928	3047
AvgHouseholdSize	2.5297	.24845	3047
PctMarriedHouseholds	51.243872141	6.5728137943	3047
PctNoHS18_24	18.224	8.0931	3047
PctHS18_24	35.002	9.0697	3047
PctBachDeg18_24	6.158	4.5291	3047
PctHS25_Over	34.805	7.0349	3047
PctBachDeg25_Over	13.282	5.3948	3047
PctUnemployed16_Over	7.852	3.4524	3047
PctPrivateCoverage	64.355	10.6471	3047
PctEmpPrivCoverage	41.196	9.4477	3047
PctPublicCoverage	36.253	7.8417	3047
PctPublicCoverageAlone	19.240	6.1130	3047
PctWhite	83.645286235	16.380025229	3047
PctBlack	9.1079776146	14.534537922	3047
PctAsian	1.2539649642	2.6102763927	3047
PctOtherRace	1.9835230038	3.5177101375	3047

- Including descriptive statistics, we can also find inferential statistics like ANOVA, R Square value. The next stage is to build a regression model using all the socioeconomic variables that are currently available.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	PctOtherRace, PctBachDeg18_24, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage[b]	.	Enter

a. Dependent Variable: deathRate

b. All requested variables entered.

- Moving on to the subsequent phase, we will construct a regression model that utilizes all of the available socioeconomic variables. The R-squared value for this model is .542, and a higher R-squared value indicates a superior fit of the regression model to the dataset

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.736[a]	.542	.538	18.8599

a. Predictors: (Constant), PctOtherRace, PctBachDeg18_24, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

- **The ANOVA (Analysis of Variance) table produces an F statistic and a P value. If the P value is below a predetermined level of significance (0.05), then we can reject the null hypothesis and conclude that a significant difference exists. In this particular dataset, the significance value is .000, which is less than 0.05. Hence, we can infer that the model is indeed significant.**

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1270601.342	23	55243.537	155.312	.000[b]
	Residual	1075264.533	3023	355.695		
	Total	2345865.875	3046			

a. Dependent Variable: deathRate

b. Predictors: (Constant), PctOtherRace, PctBachDeg18_24, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

- **We can utilize a variety of variable selection techniques that are applicable to regression analysis to obtain the best R square value. SPSS offers a variety of selection techniques, including stepwise selection, backward elimination, and forward selection.**

- **Forward Selection**

  The following variables were entered for the forward selection: incidenceRate, PctBachDeg25 Over, povertyPercent, PctOtherRace, The following statistics are for the United States: PctHS18 24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25 Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16 Over, PctPrivateCoverage, PctEmpPrivCoverage, and PctWhite.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	incidenceRate	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
2	PctBachDeg25_Over	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
3	povertyPercent	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
4	PctOtherRace	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
5	PctHS18_24	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
6	PctMarriedHouseholds	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
7	MedianAgeFemale	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
8	PctPublicCoverageAlone	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
9	PctHS25_Over	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
10	PctPublicCoverage	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
11	AvgHouseholdSize	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
12	PctUnemployed16_Over	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
13	PctPrivateCoverage	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
14	PctEmpPrivCoverage	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
15	PctWhite	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: deathRate

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.512[a]	.262	.262	23.8372
2	.683[b]	.467	.466	20.2739
3	.710[c]	.504	.503	19.5623
4	.716[d]	.513	.513	19.3754
5	.721[e]	.520	.519	19.2440
6	.724[f]	.524	.523	19.1726
7	.725[g]	.526	.525	19.1224
8	.728[h]	.529	.528	19.0635
9	.729[i]	.531	.530	19.0322
10	.730[j]	.532	.531	19.0114
11	.731[k]	.534	.532	18.9843
12	.732[l]	.536	.534	18.9494
13	.733[m]	.537	.535	18.9245
14	.734[n]	.539	.537	18.8894
15	.735[o]	.540	.538	18.8595

a. Predictors: (Constant), incidenceRate

b. Predictors: (Constant), incidenceRate, PctBachDeg25_Over

c. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent

d. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace

e. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24

f. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds

g. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale

h. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone

i. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over

j. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage

k. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize

l. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over

m. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage

n. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage

o. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage, PctWhite

- The model summary for the forward selection approach indicates that the R-squared value is .540, which is identical to the value obtained for the model that utilizes all available variables. Therefore, even when utilizing the forward selection approach, the R-squared value remains unaltered.
- Even after using forward selection, the model is significant according to the ANOVA table's Sig value of.000.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.735[a]	.540	.538	18.8595

a. Predictors: (Constant), PctOtherRace, PctMarriedHouseholds, PctBachDeg25_Over, incidenceRate, MedianAgeFemale, PctHS18_24, PctUnemployed16_Over, PctWhite, PctEmpPrivCoverage, AvgHouseholdSize, PctHS25_Over, PctPublicCoverageAlone, povertyPercent, PctPrivateCoverage, PctPublicCoverage

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1267797.812	15	84519.854	237.628	.000[b]
	Residual	1078068.063	3031	355.681		
	Total	2345865.875	3046			

a. Dependent Variable: deathRate

b. Predictors: (Constant), PctOtherRace, PctMarriedHouseholds, PctBachDeg25_Over, incidenceRate, MedianAgeFemale, PctHS18_24, PctUnemployed16_Over, PctWhite, PctEmpPrivCoverage, AvgHouseholdSize, PctHS25_Over, PctPublicCoverageAlone, povertyPercent, PctPrivateCoverage, PctPublicCoverage

- **Backward Elimination**

In Backward Elimination the variables entered and removed are:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	PctOtherRace, PctBachDeg18_24, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage[b]	.	Enter
2	.	PctBachDeg18_24	Backward (criterion: Probability of F-to-remove >= .100).
3	.	PctAsian	Backward (criterion: Probability of F-to-remove >= .100).
4	.	PctMarriedHouseholds	Backward (criterion: Probability of F-to-remove >= .100).
5	.	MedianAge	Backward (criterion: Probability of F-to-remove >= .100).
6	.	PctPublicCoverage	Backward (criterion: Probability of F-to-remove >= .100).
7	.	PctPublicCoverageAlone	Backward (criterion: Probability of F-to-remove >= .100).
8	.	medIncome	Backward (criterion: Probability of F-to-remove >= .100).
9	.	MedianAgeMale	Backward (criterion: Probability of F-to-remove >= .100).
10	.	PctBlack	Backward (criterion: Probability of F-to-remove >= .100).
11	.	Population	Backward (criterion: Probability of F-to-remove >= .100).
12	.	PctNoHS18_24	Backward (criterion: Probability of F-to-remove >= .100).

a. Dependent Variable: deathRate

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.736[a]	.542	.538	18.8599
2	.736[b]	.542	.538	18.8568
3	.736[c]	.542	.538	18.8537
4	.736[d]	.542	.539	18.8511
5	.736[e]	.542	.539	18.8484
6	.736[f]	.542	.539	18.8461
7	.736[g]	.541	.539	18.8454
8	.736[h]	.541	.539	18.8458
9	.736[i]	.541	.539	18.8468
10	.735[j]	.541	.539	18.8492
11	.735[k]	.541	.539	18.8517
12	.735[l]	.540	.538	18.8557

a. Predictors: (Constant), PctOtherRace, PctBachDeg18_24, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

b. Predictors: (Constant), PctOtherRace, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

c. Predictors: (Constant), PctOtherRace, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

d. Predictors: (Constant), PctOtherRace, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

e. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

f. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage

g. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage

h. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, MedianAgeMale, PctPrivateCoverage

i. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

j. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

k. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

l. Predictors: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, PctHS18_24, PctUnemployed16_Over, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

- The R Square value remains unchanged at .540 even after applying the backward elimination method, indicating that the model with all variables is identical to the one obtained through this method. In other words, the backward elimination method did not result in any significant change in the R square value.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.735[a]	.540	.538	18.8557

a. Predictors: (Constant), PctOtherRace, PctUnemployed16_Over, PctHS18_24, incidenceRate, MedianAgeFemale, PctBachDeg25_Over, PctWhite, PctEmpPrivCoverage, AvgHouseholdSize, PctHS25_Over, povertyPercent, PctPrivateCoverage

- Even after using backward elimination, the model is significant since the Significant value of.000 in the ANOVA table.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1267165.522	12	105597.127	297.007	.000[b]
	Residual	1078700.353	3034	355.537		
	Total	2345865.875	3046			

a. Dependent Variable: deathRate

b. Predictors: (Constant), PctOtherRace, PctUnemployed16_Over, PctHS18_24, incidenceRate, MedianAgeFemale, PctBachDeg25_Over, PctWhite, PctEmpPrivCoverage, AvgHouseholdSize, PctHS25_Over, povertyPercent, PctPrivateCoverage

- Excluded Variables are shown below:

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
2	PctBachDeg18_24	.002[b]	.130	.896	.002	.517
3	PctBachDeg18_24	.002[c]	.120	.905	.002	.520
	PctAsian	-.002[c]	-.139	.889	-.003	.515
4	PctBachDeg18_24	.001[d]	.063	.950	.001	.531
	PctAsian	-.003[d]	-.150	.881	-.003	.516
	PctMarriedHouseholds	.009[d]	.382	.703	.007	.288
5	PctBachDeg18_24	.001[e]	.065	.948	.001	.531
	PctAsian	-.002[e]	-.145	.884	-.003	.516
	PctMarriedHouseholds	.009[e]	.386	.699	.007	.288
	MedianAge	-.005[e]	-.391	.696	-.007	.978
6	PctBachDeg18_24	.002[f]	.095	.925	.002	.533
	PctAsian	-.003[f]	-.155	.877	-.003	.516
	PctMarriedHouseholds	.010[f]	.436	.663	.008	.291
	MedianAge	-.005[f]	-.384	.701	-.007	.979
	PctPublicCoverage	-.029[f]	-.484	.628	-.009	.043
7	PctBachDeg18_24	.002[g]	.137	.891	.002	.534
	PctAsian	-.002[g]	-.126	.900	-.002	.516
	PctMarriedHouseholds	.008[g]	.373	.709	.007	.293
	MedianAge	-.005[g]	-.411	.681	-.007	.979
	PctPublicCoverage	.016[g]	.489	.625	.009	.150
	PctPublicCoverageAlone	.028[g]	.890	.374	.016	.154
8	PctBachDeg18_24	.005[h]	.286	.775	.005	.545
	PctAsian	.001[h]	.070	.944	.001	.534
	PctMarriedHouseholds	.012[h]	.554	.580	.010	.302
	MedianAge	-.005[h]	-.420	.675	-.008	.980
	PctPublicCoverage	.013[h]	.416	.677	.008	.150
	PctPublicCoverageAlone	.029[h]	.917	.359	.017	.154
	medIncome	.034[h]	1.057	.290	.019	.147
9	PctBachDeg18_24	.003[i]	.210	.834	.004	.547
	PctAsian	.001[i]	.054	.957	.001	.534
	PctMarriedHouseholds	.011[i]	.511	.609	.009	.303
	MedianAge	-.006[i]	-.455	.649	-.008	.980
	PctPublicCoverage	.010[i]	.330	.741	.006	.151
	PctPublicCoverageAlone	.029[i]	.913	.361	.017	.154
	medIncome	.033[i]	1.025	.305	.019	.147
	MedianAgeMale	-.042[i]	-1.153	.249	-.021	.114
10	PctBachDeg18_24	.004[j]	.220	.826	.004	.547
	PctAsian	.008[j]	.502	.616	.009	.603
	PctMarriedHouseholds	.010[j]	.460	.646	.008	.303
	MedianAge	-.006[j]	-.452	.652	-.008	.980
	PctPublicCoverage	.016[j]	.498	.618	.009	.154
	PctPublicCoverageAlone	.032[j]	1.009	.313	.018	.155
	medIncome	.036[j]	1.120	.263	.020	.148
	MedianAgeMale	-.038[j]	-1.043	.297	-.019	.114
	PctBlack	-.035[j]	-1.335	.182	-.024	.220
11	PctBachDeg18_24	.001[k]	.086	.931	.002	.553
	PctAsian	.001[k]	.067	.947	.001	.666
	PctMarriedHouseholds	.013[k]	.575	.565	.010	.305
	MedianAge	-.005[k]	-.438	.661	-.008	.981
	PctPublicCoverage	.014[k]	.449	.653	.008	.154
	PctPublicCoverageAlone	.029[k]	.918	.359	.017	.156
	medIncome	.034[k]	1.074	.283	.020	.148
	MedianAgeMale	-.037[k]	-1.004	.316	-.018	.114
	PctBlack	-.034[k]	-1.300	.194	-.024	.220
	Population	-.018[k]	-1.337	.181	-.024	.795
12	PctBachDeg18_24	.006[l]	.361	.718	.007	.572
	PctAsian	.004[l]	.236	.814	.004	.675
	PctMarriedHouseholds	.008[l]	.355	.723	.006	.312
	MedianAge	-.005[l]	-.429	.668	-.008	.981
	PctPublicCoverage	.021[l]	.671	.502	.012	.158
	PctPublicCoverageAlone	.035[l]	1.150	.250	.021	.160
	medIncome	.028[l]	.876	.381	.016	.151
	MedianAgeMale	-.035[l]	-.961	.337	-.017	.114
	PctBlack	-.035[l]	-1.318	.188	-.024	.220
	Population	-.017[l]	-1.221	.222	-.022	.800
	PctNoHS18_24	-.023[l]	-1.515	.130	-.028	.662

a. Dependent Variable: deathRate

b. Predictors in the Model: (Constant), PctOtherRace, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

c. Predictors in the Model: (Constant), PctOtherRace, PctMarriedHouseholds, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

d. Predictors in the Model: (Constant), PctOtherRace, MedianAge, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

e. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage, PctPublicCoverage

f. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage

g. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeMale, PctPrivateCoverage

h. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, MedianAgeMale, PctPrivateCoverage

i. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctBlack, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

j. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, Population, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

k. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, PctNoHS18_24, PctHS18_24, PctUnemployed16_Over, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

l. Predictors in the Model: (Constant), PctOtherRace, incidenceRate, MedianAgeFemale, PctHS18_24, PctUnemployed16_Over, PctHS25_Over, PctEmpPrivCoverage, AvgHouseholdSize, PctBachDeg25_Over, povertyPercent, PctWhite, PctPrivateCoverage

- **Stepwise Selection:** Variables entered and removed are given below:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	incidenceRate	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	PctBachDeg25_Over	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	povertyPercent	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	PctOtherRace	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
5	PctHS18_24	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
6	PctMarriedHouseholds	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
7	MedianAgeFemale	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
8	PctPublicCoverageAlone	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
9	PctHS25_Over	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
10	PctPublicCoverage	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
11	.	MedianAgeFemale	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
12	AvgHouseholdSize	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
13	PctUnemployed16_Over	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
14	PctPrivateCoverage	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
15	PctEmpPrivCoverage	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
16	MedianAgeMale	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
17	.	PctMarriedHouseholds	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
18	.	PctPublicCoverage	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
19	.	PctPublicCoverageAlone	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
20	PctWhite	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: deathRate

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.512[a]	.262	.262	23.8372
2	.683[b]	.467	.466	20.2739
3	.710[c]	.504	.503	19.5623
4	.716[d]	.513	.513	19.3754
5	.721[e]	.520	.519	19.2440
6	.724[f]	.524	.523	19.1726
7	.725[g]	.526	.525	19.1224
8	.728[h]	.529	.528	19.0635
9	.729[i]	.531	.530	19.0322
10	.730[j]	.532	.531	19.0114
11	.729[k]	.532	.531	19.0111
12	.730[l]	.533	.532	18.9930
13	.731[m]	.535	.533	18.9599
14	.732[n]	.536	.534	18.9442
15	.733[o]	.537	.535	18.9145
16	.734[p]	.539	.537	18.8883
17	.734[q]	.539	.537	18.8907
18	.734[r]	.538	.537	18.8931
19	.734[s]	.538	.537	18.8919
20	.735[t]	.540	.538	18.8638

a. Predictors: (Constant), incidenceRate

b. Predictors: (Constant), incidenceRate, PctBachDeg25_Over

c. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent

d. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace

e. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24

f. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds

g. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale

h. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone

i. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over

j. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, MedianAgeFemale, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage

k. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage

l. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize

m. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over

n. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage

o. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage

p. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctMarriedHouseholds, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage, MedianAgeMale

q. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctPublicCoverageAlone, PctHS25_Over, PctPublicCoverage, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage, MedianAgeMale

r. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctPublicCoverageAlone, PctHS25_Over, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage, MedianAgeMale

s. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctHS25_Over, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage, MedianAgeMale

t. Predictors: (Constant), incidenceRate, PctBachDeg25_Over, povertyPercent, PctOtherRace, PctHS18_24, PctHS25_Over, AvgHouseholdSize, PctUnemployed16_Over, PctPrivateCoverage, PctEmpPrivCoverage, MedianAgeMale, PctWhite

- The model summary for the stepwise selection approach reveals that the R Square value is .540, which is identical to the value obtained for the model that includes all the variables. Thus, even after applying the stepwise selection procedure, there is no change in the R square value.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.735[a]	.540	.538	18.8638

a. Predictors: (Constant), PctOtherRace, PctUnemployed16_Over, PctHS18_24, incidenceRate, MedianAgeMale, PctBachDeg25_Over, PctWhite, AvgHouseholdSize, PctEmpPrivCoverage, PctHS25_Over, povertyPercent, PctPrivateCoverage

- Even after using stepwise selection, the model is significant since the ANOVA table's Sig value of.000.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1266237.772	12	105519.814	296.535	.000[b]
	Residual	1079628.103	3034	355.843		
	Total	2345865.875	3046			

a. Dependent Variable: deathRate

b. Predictors: (Constant), PctOtherRace, PctUnemployed16_Over, PctHS18_24, incidenceRate, MedianAgeMale, PctBachDeg25_Over, PctWhite, AvgHouseholdSize, PctEmpPrivCoverage, PctHS25_Over, povertyPercent, PctPrivateCoverage

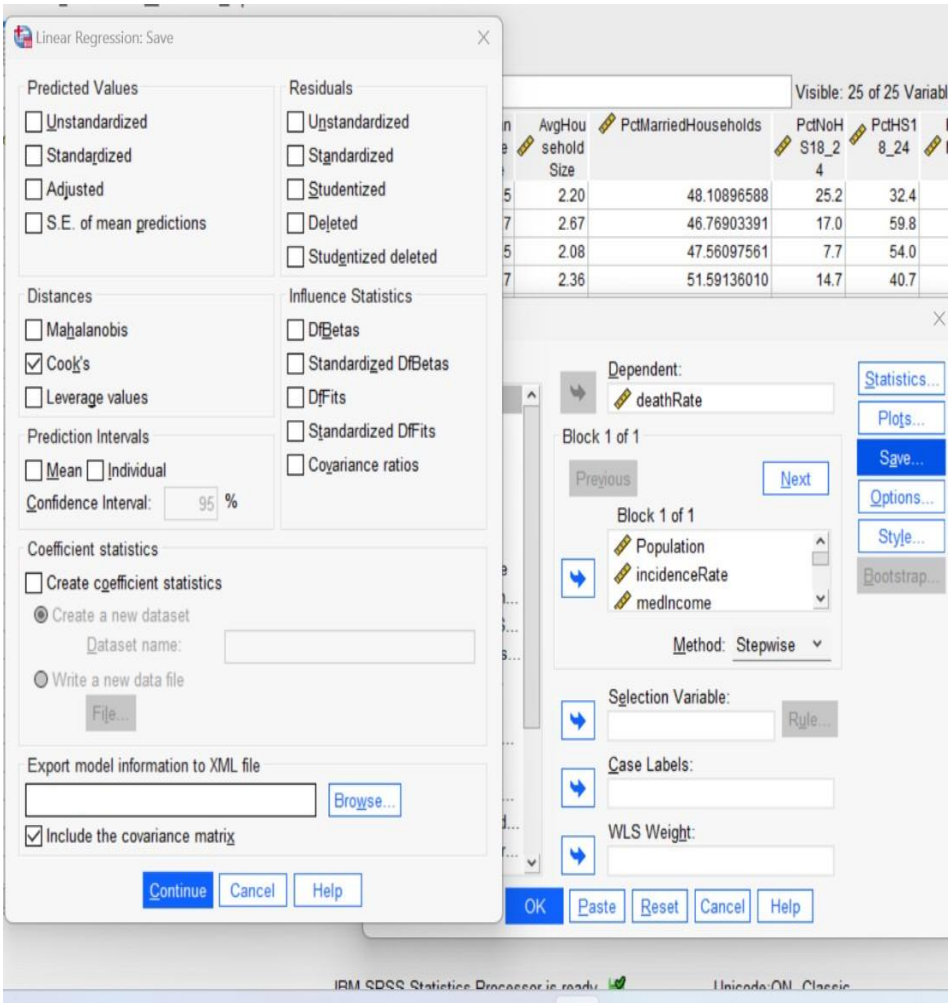
The R Square value of .540 is consistently obtained across three different variable selection methods (forward, reverse, and stepwise), indicating that the elimination of extraneous variables did not significantly affect the R square value. In other words, the R square value remains unchanged despite using different variable selection methods.

➢ Outlier Removal

Cook's Method

By detecting observations with large Cook's distances, outliers can be removed from the dataset using Cook's approach.

- We must choose the Cook's Distance from Save option in SPSS.



- A new variable called COO_1 is established, indicating the cook's distance.

PctPublicCoverage	PctPublicCoverageAlo..	PctWhite	PctBlack	PctAsian	PctOtherRace	COO_1
53.2	23.4	96.14582103	.625959608	.519664580	.141726704	.00362
53.8	27.1	98.44266163	.070788108	.165172251	.141576215	.00257
48.3	25.6	90.28309105	.612088753	.816118337	2.384595766	.01202
36.2	18.6	74.84473039	21.417651370	.342389400	1.576583750	.00005
42.7	26.9	81.67343116	7.461754605	.312207306	6.244146113	.01165
29.2	11.2	98.32735962	.358422939	.000000000	.000000000	.00001
37.0	16.0	95.23966555	2.855136906	.346786375	.371252899	.00103
37.4	18.6	96.60470880	.570012392	.000000000	.470879802	.00859
27.3	11.2	93.41481745	1.467505241	1.791068391	.798227918	.00004
35.1	17.5	94.15519059	2.902035339	.754065885	.270792728	.00066
31.1	11.9	97.95643661	.838228096	.189672051	.140724425	.00073
39.3	22.3	68.76623707	27.377617030	.430441648	1.533289185	.00035
36.2	17.0	92.21624756	5.071432337	.218778006	.548262955	.00009
42.4	25.4	96.91872792	1.469964664	.000000000	.098939929	.01263
36.4	17.3	58.77102643	31.536789110	.125872525	5.675706603	.00087
24.4	9.3	96.81388013	.954258675	.563880126	.339116719	.00344
31.3	15.5	95.08368800	1.131744184	.345439624	1.633681318	.00001
42.8	25.6	59.15456929	35.160196970	.511607086	3.696361195	.01040

- To improve the fit of the regression model, a new column called COO_1 is added that contains the Cook's distance values. Any value exceeding 0.00131 is eliminated from the dataset. Each observation is then compared to this value. This process helps to refine the regression model and improve its accuracy..
- Sorting the table in ascending order based on the Cook's distance data enables us to easily eliminate variables and delete values that fall below a specified Cook's distance. This makes the process of removing variables from the dataset more straightforward and efficient.
- After eliminating the variables, we proceed to develop a regression model.
- Below is a list of values that have been either added or removed after implementing

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Cook's Distance, PctHS25_Over, MedianAge, PctMarriedHouseholds, PctNoHS18_24, incidenceRate, AvgHouseholdSize, Population, PctHS18_24, PctOtherRace, PctBachDeg18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctEmpPrivCoverage, MedianAgeMale, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeFemale, PctPrivateCoverage, PctPublicCoverage[b]	.	Enter

a. Dependent Variable: deathRate

b. All requested variables entered.

the Cook's approach.

- The removal of the outlier using Cook's approach resulted in a significant increase in the R square value, from .540 to .640. This improvement indicates that the model is now the best fit for the dataset.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.800[a]	.640	.637	14.9684

a. Predictors: (Constant), Cook's Distance, PctHS25_Over, MedianAge, PctMarriedHouseholds, PctNoHS18_24, incidenceRate, AvgHouseholdSize, Population, PctHS18_24, PctOtherRace, PctBachDeg18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctEmpPrivCoverage, MedianAgeMale, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeFemale, PctPrivateCoverage, PctPublicCoverage

- The application of Cook's approach to remove outliers resulted in a Sig value of .000 in the ANOVA table, which indicates that the model is still significant even after removing the outliers.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1119083.363	24	46628.473	208.115	.000[b]
	Residual	629809.519	2811	224.052		
	Total	1748892.882	2835			

a. Dependent Variable: deathRate

b. Predictors: (Constant), Cook's Distance, PctHS25_Over, MedianAge, PctMarriedHouseholds, PctNoHS18_24, incidenceRate, AvgHouseholdSize, Population, PctHS18_24, PctOtherRace, PctBachDeg18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctEmpPrivCoverage, MedianAgeMale, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeFemale, PctPrivateCoverage, PctPublicCoverage

- This is a coefficient table for the data using Cook's method:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	194.009	13.733		14.128	<.001
	Population	-8.474E-7	.000	-.012	-.842	.400
	incidenceRate	.195	.006	.416	33.034	<.001
	medIncome	.000	.000	.059	1.846	.065
	povertyPercent	.497	.129	.120	3.856	<.001
	MedianAge	-.003	.007	-.006	-.521	.602
	MedianAgeMale	-.442	.172	-.090	-2.574	.010
	MedianAgeFemale	-.136	.186	-.028	-.730	.465
	AvgHouseholdSize	-16.853	2.563	-.153	-6.577	<.001
	PctMarriedHouseholds	.024	.085	.006	.280	.780
	PctNoHS18_24	-.111	.046	-.034	-2.389	.017
	PctHS18_24	.260	.042	.091	6.256	<.001
	PctBachDeg18_24	.051	.092	.009	.555	.579
	PctHS25_Over	.348	.080	.098	4.345	<.001
	PctBachDeg25_Over	-1.250	.127	-.269	-9.860	<.001
	PctUnemployed16_Over	.538	.137	.071	3.937	<.001
	PctPrivateCoverage	-.947	.115	-.383	-8.205	<.001
	PctEmpPrivCoverage	.483	.088	.179	5.483	<.001
	PctPublicCoverage	.011	.181	.003	.059	.953
	PctPublicCoverageAlone	-.211	.230	-.049	-.919	.358
	PctWhite	-.133	.056	-.083	-2.383	.017
	PctBlack	-.043	.053	-.024	-.806	.420
	PctAsian	-.075	.163	-.008	-.459	.646
	PctOtherRace	-1.139	.121	-.140	-9.414	<.001
	Cook's Distance	-543.913	1082.324	-.006	-.503	.615

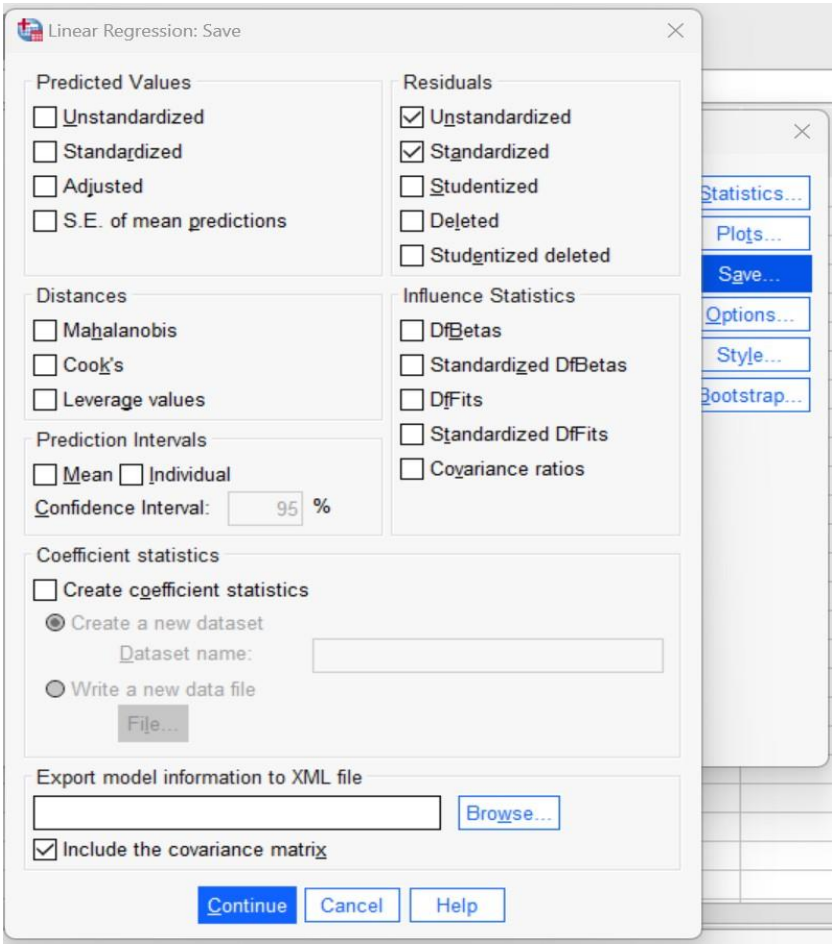
a. Dependent Variable: deathRate

• Assumptions

- The significance of the model is confirmed by the Sig value of .000 obtained from the ANOVA table after removing the outliers using Cook's approach. To evaluate the precision of the model and verify its results, it is essential to consider the underlying assumptions of linear regression. These assumptions include:

➢ Linearity assumption

In order to assess the accuracy of the model and validate its results, it is necessary to have a

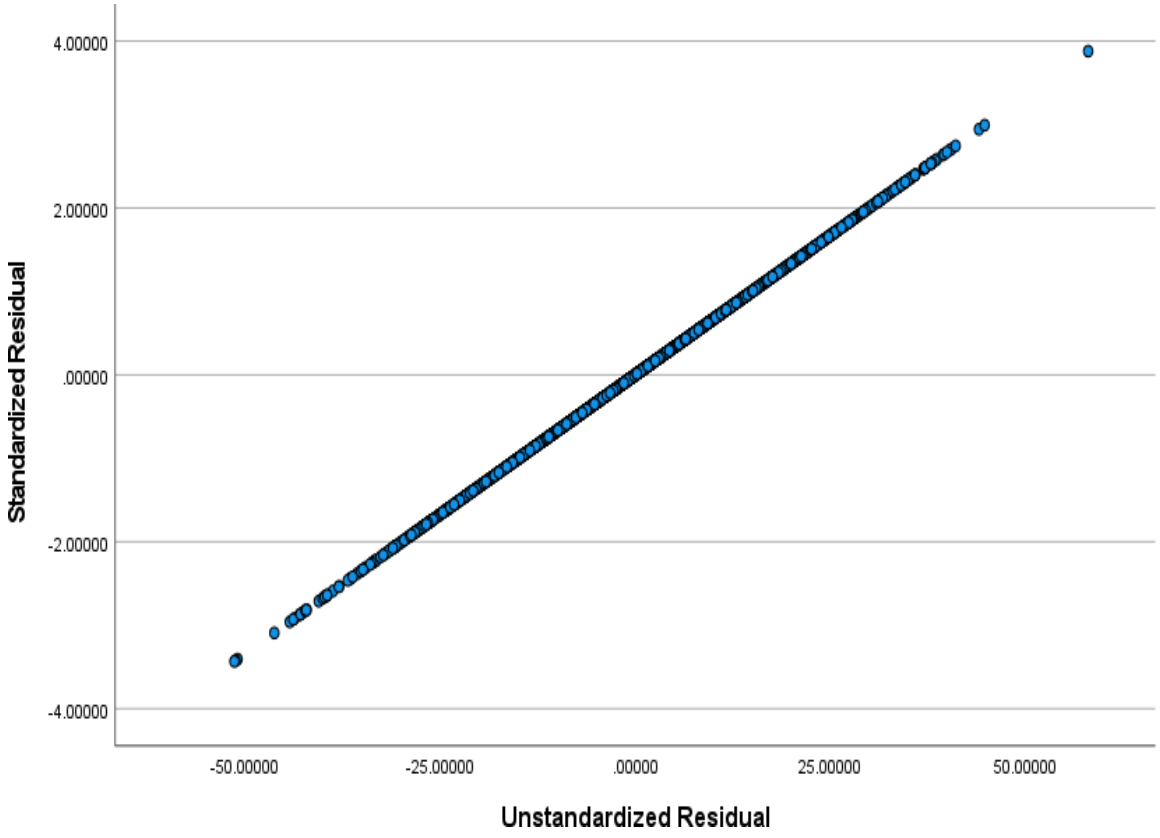


clear understanding of the underlying assumptions of linear regression. Once the analysis is complete, we can select the Standardized and Unstandardized options from the Save menu in SPSS..

- It creates two variables in the data sheet named:
  RES – Unstandardised
  ZRES - Standardised

RES_1	ZRE_1
-.12061	-.00806
-2.84948	-.19037
.38460	.02569
1.72919	.11552
-.56616	-.03782
.65089	.04348
.18006	.01203
-.36137	-.02414
-.79658	-.05322
.72955	.04874
-3.37961	-.22578
.41755	.02790
-.51969	-.03472
-.24812	-.01658
2.05124	.13704
.61733	.04124
1.07100	.07155
.07636	.00510

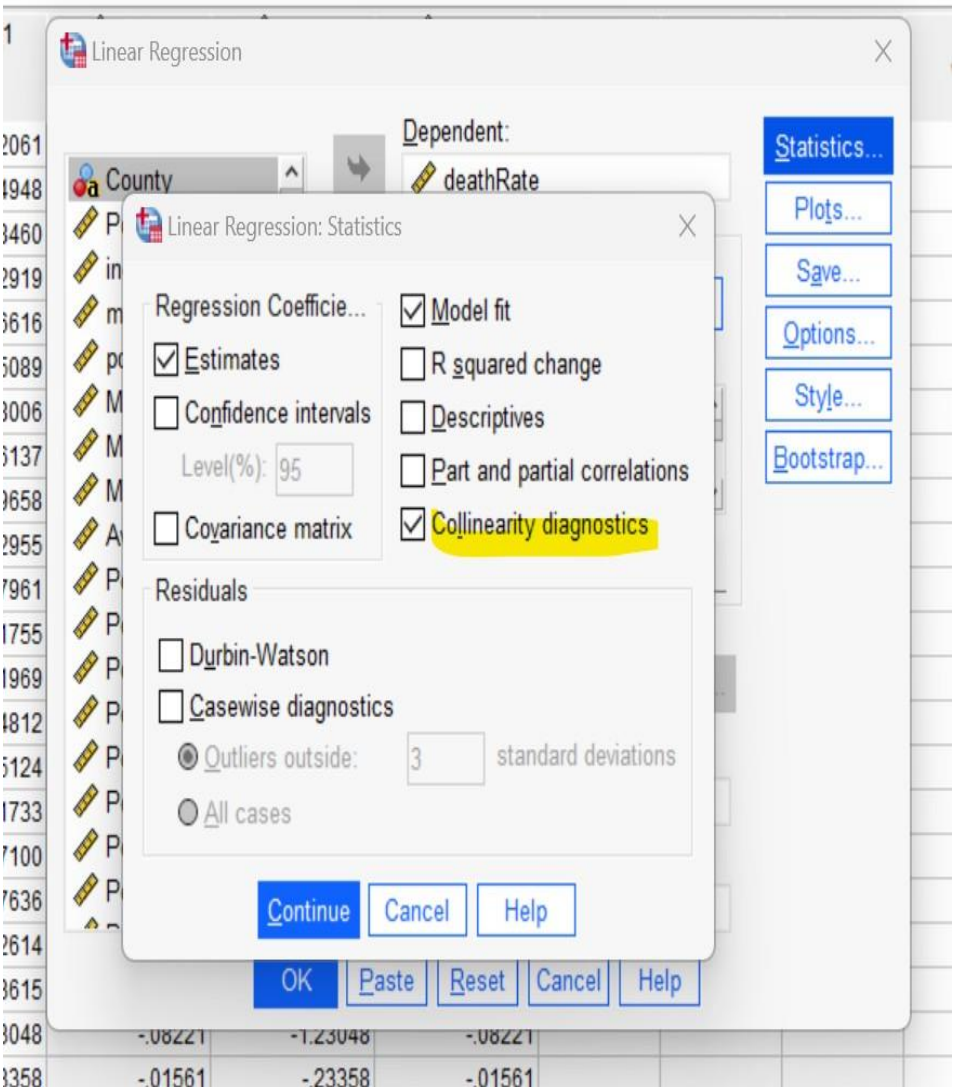
- The following step is to create a graph to see if the relationship between the variables is linear. In SPSS, we choose a scatterplot from the graph.
- **X axis will be Unstandardized and Y axis will be Standardised variable**.



As all of the points are on a straight line and the relationship between the dependent and independent variables is linear, the model is said to be based on the linear assumption.

- **No Multi collinearity**

- Rerun the model in SPSS and select the statistics option for collinearity diagnostics.



- Once the Collinearity Diagnostics option is selected, various tables are generated in the Output box, including the ANOVA, Model Summary, Coefficients, Collinearity Diagnostics, and Residual Statistics tables.
- The Coefficients table indicates that most of the collinearity statistics values are greater than 5, which suggests that the model is experiencing multicollinearity. This finding contradicts our assumption that there is no multicollinearity present in the model.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	194.009	13.733		14.128	<.001		
	Population	-8.474E-7	.000	-.012	-.842	.400	.675	1.480
	incidenceRate	.195	.006	.416	33.034	<.001	.809	1.237
	medIncome	.000	.000	.059	1.846	.065	.127	7.876
	povertyPercent	.497	.129	.120	3.856	<.001	.132	7.551
	MedianAge	-.003	.007	-.006	-.521	.602	.979	1.021
	MedianAgeMale	-.442	.172	-.090	-2.574	.010	.106	9.472
	MedianAgeFemale	-.136	.186	-.028	-.730	.465	.088	11.385
	AvgHouseholdSize	-16.853	2.563	-.153	-6.577	<.001	.238	4.200
	PctMarriedHouseholds	.024	.085	.006	.280	.780	.268	3.737
	PctNoHS18_24	-.111	.046	-.034	-2.389	.017	.621	1.609
	PctHS18_24	.260	.042	.091	6.256	<.001	.600	1.666
	PctBachDeg18_24	.051	.092	.009	.555	.579	.493	2.029
	PctHS25_Over	.348	.080	.098	4.345	<.001	.252	3.966
	PctBachDeg25_Over	-1.250	.127	-.269	-9.860	<.001	.172	5.800
	PctUnemployed16_Over	.538	.137	.071	3.937	<.001	.398	2.512
	PctPrivateCoverage	-.947	.115	-.383	-8.205	<.001	.059	17.030
	PctEmpPrivCoverage	.483	.088	.179	5.483	<.001	.121	8.272
	PctPublicCoverage	.011	.181	.003	.059	.953	.041	24.201
	PctPublicCoverageAlone	-.211	.230	-.049	-.919	.358	.045	22.263
	PctWhite	-.133	.056	-.083	-2.383	.017	.105	9.485
	PctBlack	-.043	.053	-.024	-.806	.420	.141	7.076
	PctAsian	-.075	.163	-.008	-.459	.646	.454	2.202
	PctOtherRace	-1.139	.121	-.140	-9.414	<.001	.580	1.723
	Cook's Distance	-543.913	1082.324	-.006	-.503	.615	.922	1.085

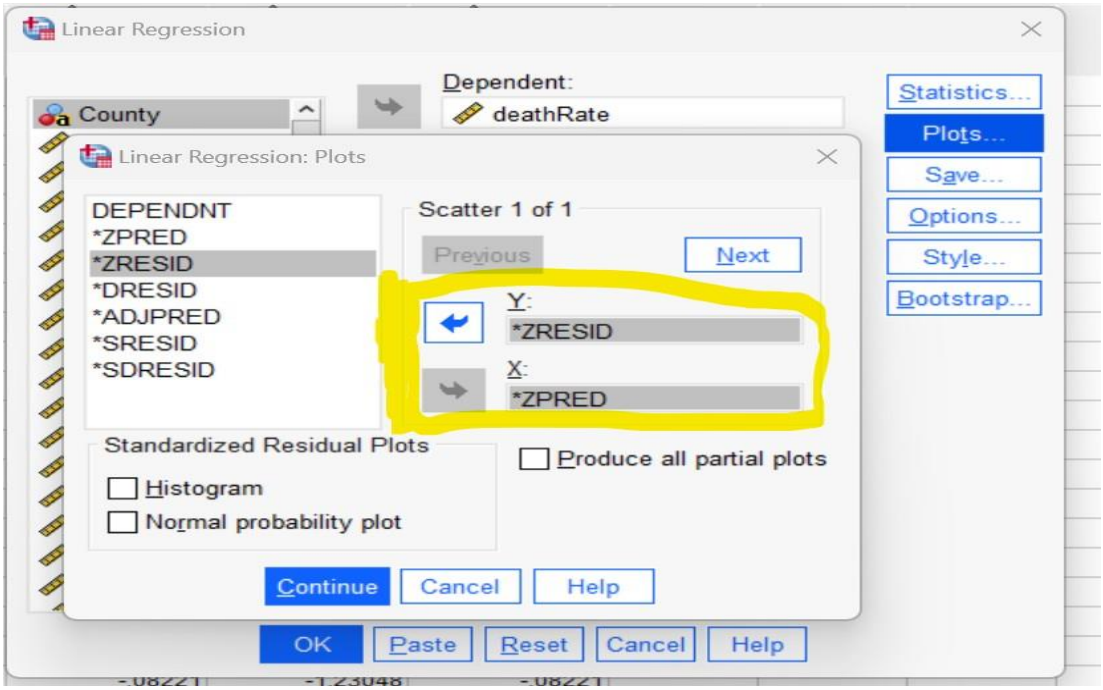
a. Dependent Variable: deathRate

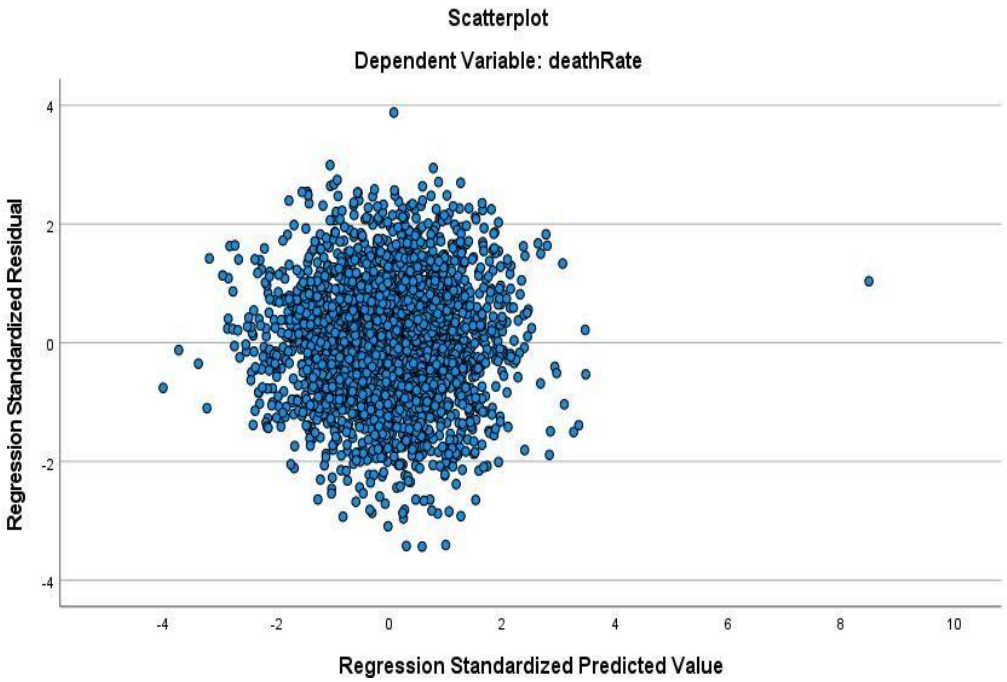
- Independent variables such as **medIncome, povertypercent, MedianAgeMale,MedianAgeFemale,PctBachDeg25_Over,PctPrivate Coverage,PctEmpPrivCoverage,PctPublicCoverage,PctPublicCoverag eAlone,PctWhite,PctBlack** are having value above 5.

As a result, this model does not adhere to the No Multi Collinearity assumption.

➢ Homoscedasticity

- In SPSS, Rerun the regression, select plots, and move ZRESID in Y axis and ZPRED in X axis.

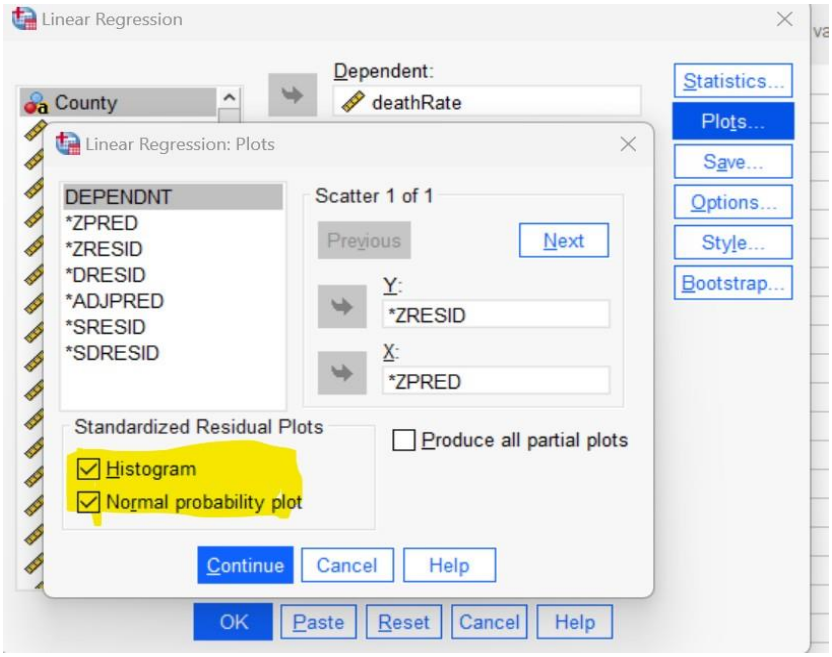


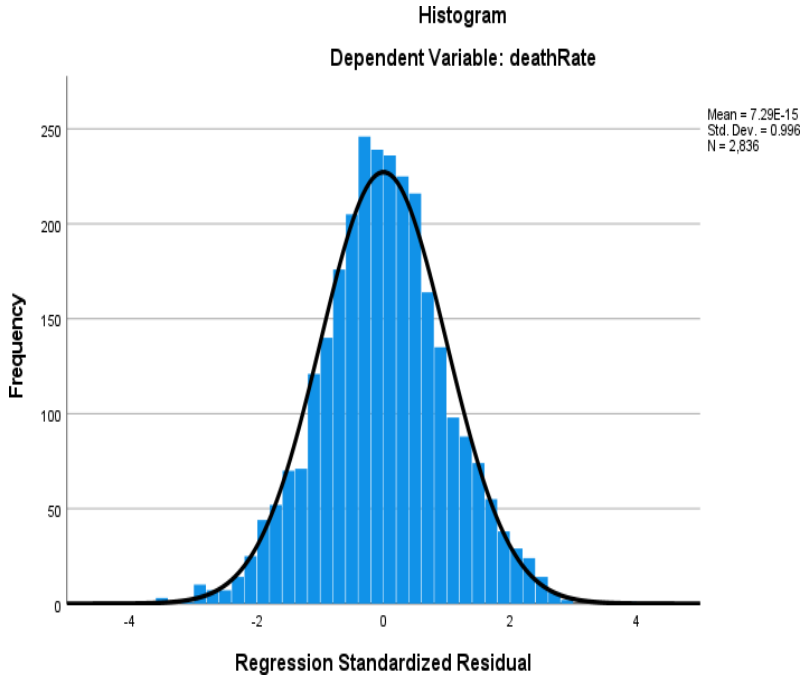


- **This model follows Homoscedasticity assumption** as all the variables are scattered around the plot.

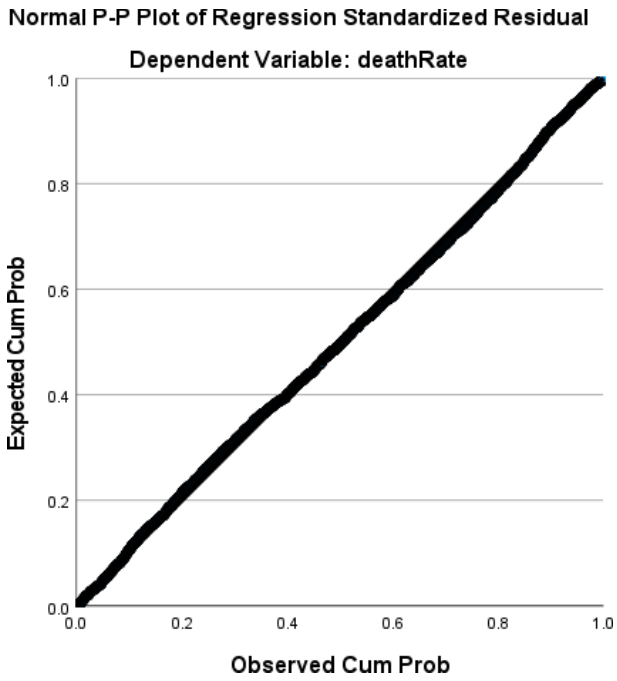
➢ Normality

- Rerun the regression in SPSS, choose the plots, tick the Histogram and Normal Probability plot boxes from the standardized Residual plot.





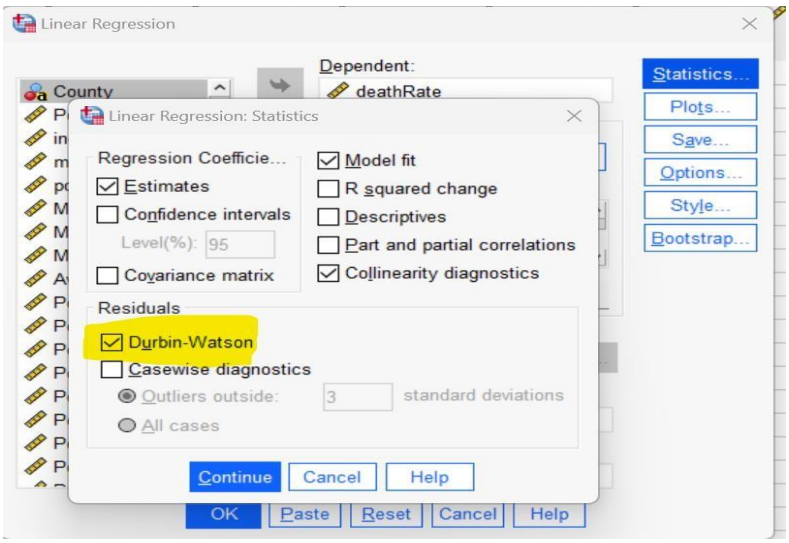
- The residuals are seen to have a normal distribution in the histogram.



- The Normal P-P Plot shows that the residuals do indeed follow a normal distribution.
- The model is using the normality assumption because both the chart have a normal distribution.

➢ No Autocorrelation

- In SPSS, rerun the regression and under statistics choose Durbin-Watson from Residuals.



Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Cook's Distance, PctHS25_Over, MedianAge, PctMarriedHouseholds, PctNoHS18_24, incidenceRate, AvgHouseholdSize, Population, PctHS18_24, PctOtherRace, PctBachDeg18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctEmpPrivCoverage, MedianAgeMale, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeFemale, PctPrivateCoverage, PctPublicCoverage[b]	.	Enter

a. Dependent Variable: deathRate

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.800[a]	.640	.637	14.9684	1.965

a. Predictors: (Constant), Cook's Distance, PctHS25_Over, MedianAge, PctMarriedHouseholds, PctNoHS18_24, incidenceRate, AvgHouseholdSize, Population, PctHS18_24, PctOtherRace, PctBachDeg18_24, PctAsian, PctUnemployed16_Over, PctBlack, PctEmpPrivCoverage, MedianAgeMale, PctPublicCoverageAlone, PctBachDeg25_Over, povertyPercent, medIncome, PctWhite, MedianAgeFemale, PctPrivateCoverage, PctPublicCoverage

b. Dependent Variable: deathRate

The Durbin-Watson statistic value, which is close to 2, suggests that the model does not exhibit autocorrelation in the data. The model satisfies all four assumptions, including Linearity, Normality, No Autocorrelation, and Homoscedasticity, except for No Multicollinearity, as there is a high level of multicollinearity among most of the independent variables. Based on the study's results, the best technique for predicting the dependent variables using relevant independent variables is multiple linear regression, as it yields the highest R square value of 0.640.

- THANKS