Given data is life expectancy of country and other varies parameter of a country at a particular year

Linear Regression

IDV – Schooling

DV – life expectancy

R2 score = 0.5052

This means that about 50.5% of the predicted variable – life expectancy is explained by the independent variable i.e. schooling.

Intercept = 41.22 (the unexplained part of the model)

This is not a good model since r2 score is very less, hence has large portion of the prediction that is unexplained , so hence the predicted value will vary more from actual value resulting in larger inaccuracy like



Where predicted is 56 years life expectancy but the persons actual life expectancy is 48.6 years

So to improve this we go for multiple regression where the predicted and actual values deviation is less , i.e we aim for more explanatory power of model i.e more r2 score

Multiple regression

We select the independent variables based on their relationship to the dependent variable, out of the 22 variables, the 5 variables that have the most positive or negative effect on life expectancy are Schooling, Income composition of resources, BMI, HIV/AIDS, Adult mortality, we find this out using correlation matrix.

Schooling – it is highly positive correlated, this means that based on the schooling a person experiences, his life expectancy can be predicted roughly. Overall, it is one the key components that affect the life expectancy positively when a person studies for more years i.e. a highly educated person is likely to have a higher life expectancy than someone who went till only middle school for education.

ICOR (Income composition of resources): It is highly positively correlated, This mostly varies based on the country the person is from, for example people from afg have a lower icor lesser than 0.5, whereas people from Albinia have icor higher than 0.7, icor varies between 0 and 1 , the higher the icor value , the more better is the utilization of available resources. Developed economy generally have higher icor , icor also plays key part in a person life expectancy , since it dicates the living environment of their country and place they stay in that country.

BMI – bmi is unit which is too bad to have it low or high, bmi is moderately positively correlated , i.e the lower the bmi the lower the life expectancy , it is not highly positive correlated since too much high bmi also will result in lower life expectancy. Hence its good to have it balanced or slightly towards higher side.

HIV/AIDS: it is moderately negatively correlated, since a country with higher rates of HIV /aids will have lower life expectancy.

Adult mortality: it could mean the no of deaths of adults in a year in a country , higher the no of deaths , it indicates lesser the life expectancy of people in country , it would have better to convert this adult mortality death no to adult morality percentage per country based on the no oif deaths and the number of adults in the country during that year (i.e adult population census of country). This is highly negatively correlated with life expectancy as seen in correlation matrix.

Based on these reasons and correlation value , these 5 are chosen as the independent variables in order to predict the life expectancy of a country at a particular year

R2 score = 0.7959, this indicates that 79.59% of the changes in the predicated variable life expectancy is explainable by the 5 IDVs.

The explanatory power of model is satisfactory since its about 80%, but still it can be improved by further analysis and conversion of adult morality into adult mortality rate based on population.



The prediction power is higher as since here as the gap between the actual and predicted variable has decreased , also for some entries the prediction is more accurate

