**Problem 3**

**Part 1**

* We are tasked with building the best correlation to predict the permeability of a well , from the GR, DT, MSFL, DT, RHOB, NPHI, and PEF well logs.
* The dependent variable is taken as ln(Kg) and the independent variables are GR, DT, MSFL, DT, RHOB, NPHI and PEF.
* The data file ‘scfu5wells.dat’ is modified to create a new data file that only contains the above parameters for the lithotypes of 7 and 8.
* The data file is read into the GRACE application, and a variable selection procedure is applied to obtain the selection of variables that results in maximal correlation.

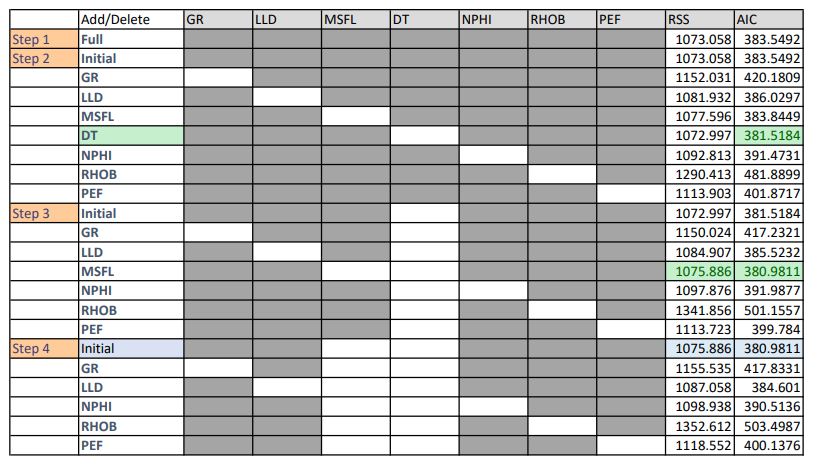


Figure :Variable selection procedure

**Step 1:-** The correlations for the initial case were generated by building a correlation for ln(Kg) with the well log data from GR, DT, log10(LLD), log10(MSFL), DT, NPHI, RHOB and PEF.(**Fig. 1**)

**Step 2:-** The initial case was the same as the case from step 1.Each of the variables were eliminated individually in Step 2, and the correlations were built for them respectively. The RSS and AIC values are recorded. The case with the lowest AIC value in step 2 was the case where the variable ‘DT’ was eliminated. The AIC value was **381.5184**.

**Step 3:**- The initial case for step 3 was the same as the case for the lowest AIC value in step 2(eliminated variable DT). Variable DT is now not to be considered for the variable selection procedure. The correlations were built by individually eliminating the variables respectively – GR, log10(LLD), log10(MSFL),NPHI, RHOB and PEF. The AIC values were recorded again. The lowest AIC value was **380.9811,**  the variable log10(MSFL) was eliminated. The least AIC value for step 3 was at **380.9811** by eliminating MSFL and DT.

**Step 4:-**  The initial case for step 4 was taken as the case with the lowest AIC(380.9811) in step 3, which is the case where log10(MSFL) and DT were eliminated. The step-wise variable selection procedure was carried out similarly to step 3. The individual variables GR, log10(LLD), NPHI, RHOB, PEF were considered in step 4. Other than the initial case, the lowest AIC recorded in step 4 was AIC of **384.601** when we eliminated log10(LLD), log10(MSFL) and DT, but this AIC was still higher than the AIC of the initial case **380.9811.** Hence, the best correlation is said to be the case where the correlations for ln(kg) are built by considering just GR, log10(LLD), NPHI, RHOB, PEF and by eliminating log10(MSFL) and DT, where the **lowest AIC** was **380.9811.**

Figure : Best fit equations for correlation built by eliminating DT and log10(MSFL). AIC = 380.9811

**Part 2**

* The data for the blind well was read into an excel sheet(Part 3.xlsx), and the correlations built in the previous part were used to generate transformed data for the different variables in the blind well. The permeability was predicted, and the predicted ln(kg) values were then compared to the measured ln(kg) values. The values are included in sheet2 of the Part3.xlsx file.

**Part 3**

* The values of the measured ln(kg) values are plotted against the predicted ln(kg) values in **Fig. 3.**

Figure : ln(k) vs ln(kg)\_Tr

**Part 4**

* A plot between the ln(kg)\_meas and the ln(kg)\_Pred was graphed. The predicted values of the ln(kg) are visible are scatterpoints, and a trendline was generated.
* The predicted values of ln(kg)\_pred can be seen to follow a linearly increasing relationship with the ln(kg)\_meas values. This shows that the correlation built by us using a variable selection procedure was successful in predicting the permeability of the blind well.
* The trendline equation was a seen to be a linear equation of the form y = mx+b.
* The correlation coefficient was 0.5324. which tells us that more than 50% of the predicted data can fit with the ln(kg)\_meas values along a linear relationship.