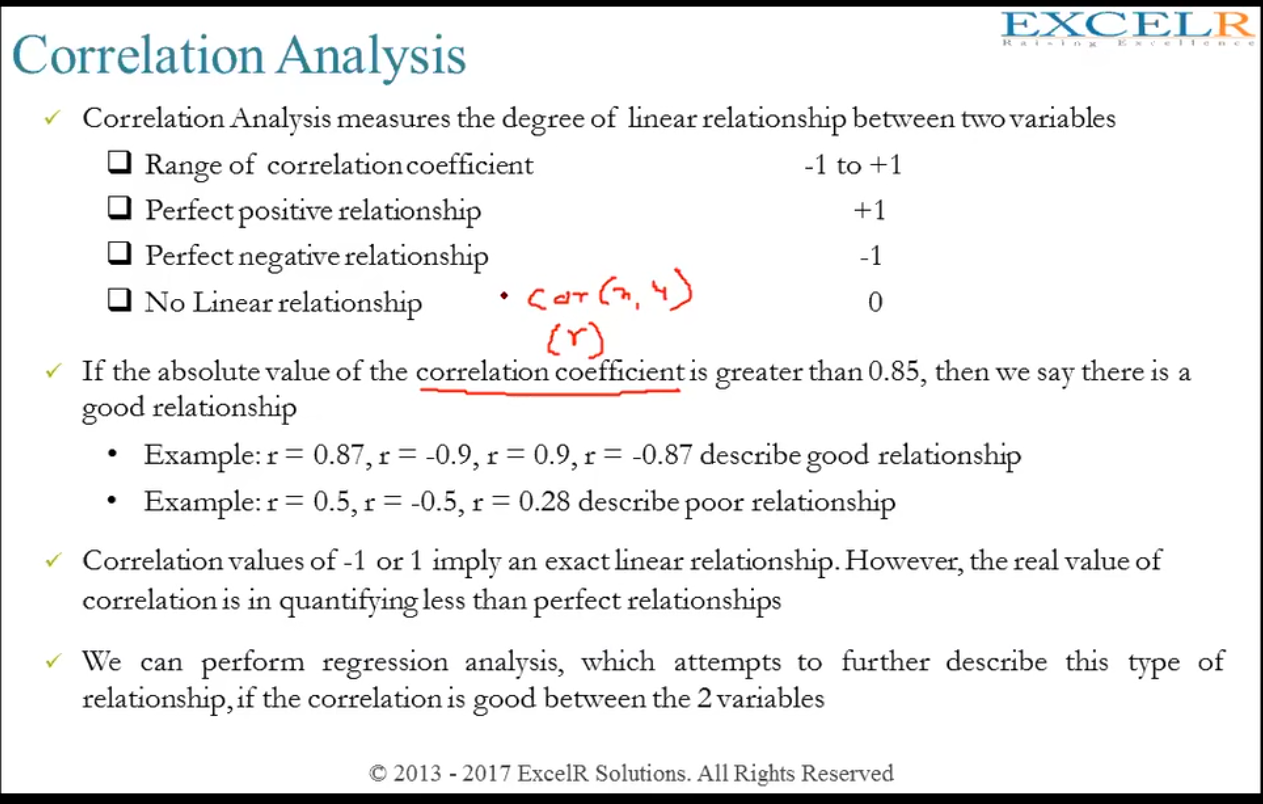
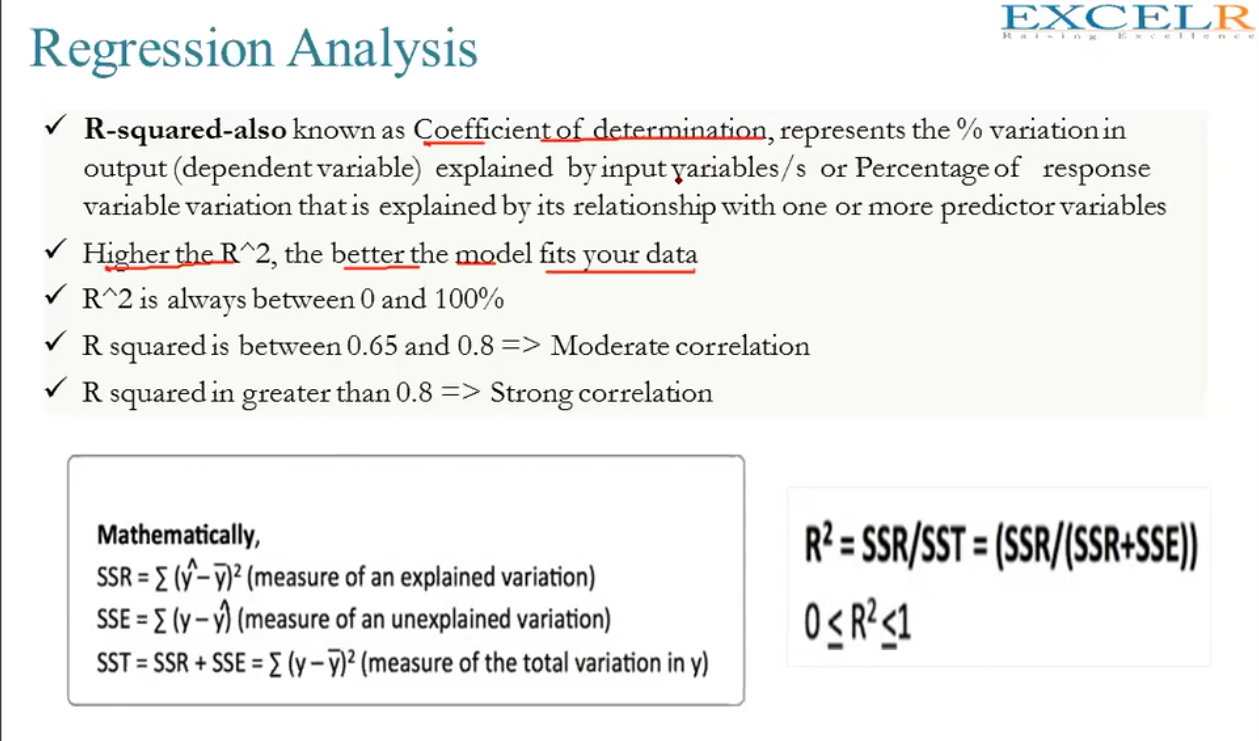
Linear regression

correlation 0.85 is value to be known



Rsqaure thing

note- there is other r square in small which is the rotte mean square error value



***Difference between ols and sklearn linear regression***

To summarize some key differences:

· OLS efficiency: scikit-learn is faster at linear regression; the difference is more apparent for larger datasets

· Logistic regression efficiency: employing only a single core, statsmodels is faster at logistic regression

· Visualization: statsmodels provides a summary table

· Solvers/ methods: in general, statsmodels provides a greater variety

· Logistic Regression: scikit-learn regularizes by default while statsmodels does not

· Additional linear models: scikit-learn provides more models for regularization, while statsmodels helps correct for broken OLS assumptions

In general, scikit-learn is designed for machine-learning, while statsmodels is made for rigorous statistics. Both libraries have their uses. Before selecting one over the other, it is best to consider the purpose of the model. A model designed for prediction is best fit using scikit-learn, while statsmodels is best employed for explanatory models. To completely disregard one for the other would do a great disservice to an excellent Python library.

***What is R-squared and adjusted r squared value??***

Let us first understand what is R-squared:

R-squared or R2 explains the degree to which your input variables explain the variation of your output / predicted variable. So, if R-square is 0.8, it means 80% of the variation in the output variable is explained by the input variables. So, in simple terms, higher the R squared, the more variation is explained by your input variables and hence better is your model.

However, the problem with R-squared is that it will either stay the same or increase with addition of more variables, even if they do not have any relationship with the output variables. This is where “Adjusted R square” comes to help. Adjusted R-square penalizes you for adding variables which do not improve your existing model.

Hence, if you are building Linear regression on multiple variable, it is always suggested that you use Adjusted R-squared to judge goodness of model. In case you only have one input variable, R-square and Adjusted R squared would be exactly same.

Typically, the more non-significant variables you add into the model, the gap in R-squared and Adjusted R-squared increases.Let us first understand what is R-squared:

R-squared or R2 explains the degree to which your input variables explain the variation of your output / predicted variable. So, if R-square is 0.8, it means 80% of the variation in the output variable is explained by the input variables. So, in simple terms, higher the R squared, the more variation is explained by your input variables and hence better is your model.

However, the problem with R-squared is that it will either stay the same or increase with addition of more variables, even if they do not have any relationship with the output variables. This is where “Adjusted R square” comes to help. Adjusted R-square penalizes you for adding variables which do not improve your existing model.

Hence, if you are building Linear regression on multiple variable, it is always suggested that you use Adjusted R-squared to judge goodness of model. In case you only have one input variable, R-square and Adjusted R squared would be exactly same.

Typically, the more non-significant variables you add into the model, the gap in R-squared and Adjusted R-squared increases.