

2. Explain the Anscombe's quartet in detail?

Answer -->

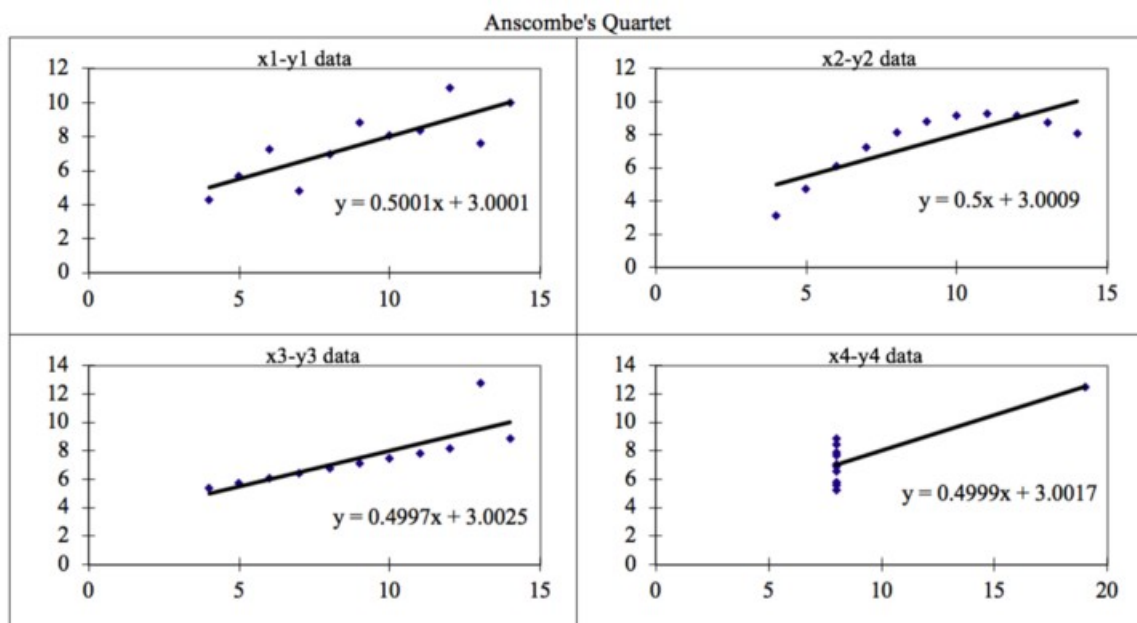
Anscombe's quartet is a set of 4 graphs that advises to look on our data before building and running linear regression models.

Linear regression models have some features that may create problems:

1. sensitiveness to outliers.
2. build linear relationship only.
3. pre-acknowledges all assumptions are being held true.

Sometimes for different types of dataset, the statistic values of the best fit line provided by the linear regression is almost the same, irrespective of the nature of values.

This problem is well defined with Anscombe's Quartet.



Explanations for graphs:

1. linear regression model is perfectly aligned with the dataset as the plot represents linear relationships.
2. though the relationship is non-linear, model still creates a linear relationship.
3. due to presence of some outliers, the best fit line is deviated from the best possible alignment, if these outliers were absent, our best fit line would be passing from all the data points.
4. few distance outliers and values concentrated over a single points may creates a best fit line which is totally false in nature.