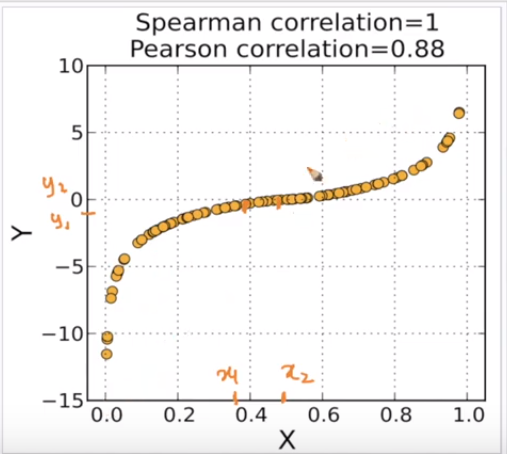
Sometimes there doesn’t exist a marked linear relationship between two random variables but a **monotonic relation** (if one increases, the other also increases or instead, decreases) is clearly noticed. A Pearson’s Correlation Coefficient evaluation, in this case, would give us the strength and direction of the linear association only between the variables of interest. Herein comes the advantage of the Spearman Rank Correlation methods, which will instead, give us the strength and direction of the monotonic relation between the connected variables.

The Spearman’s Correlation Coefficient, represented by ρ or by rR, is a nonparametric measure of the strength and direction of the association that exists between two ranked variables. It determines the degree to which a relationship is monotonic, i.e., whether there is a monotonic component of the association between two continuous or ordered variables. Below figure the curve with not a linear relations, but with a monotonic relation.

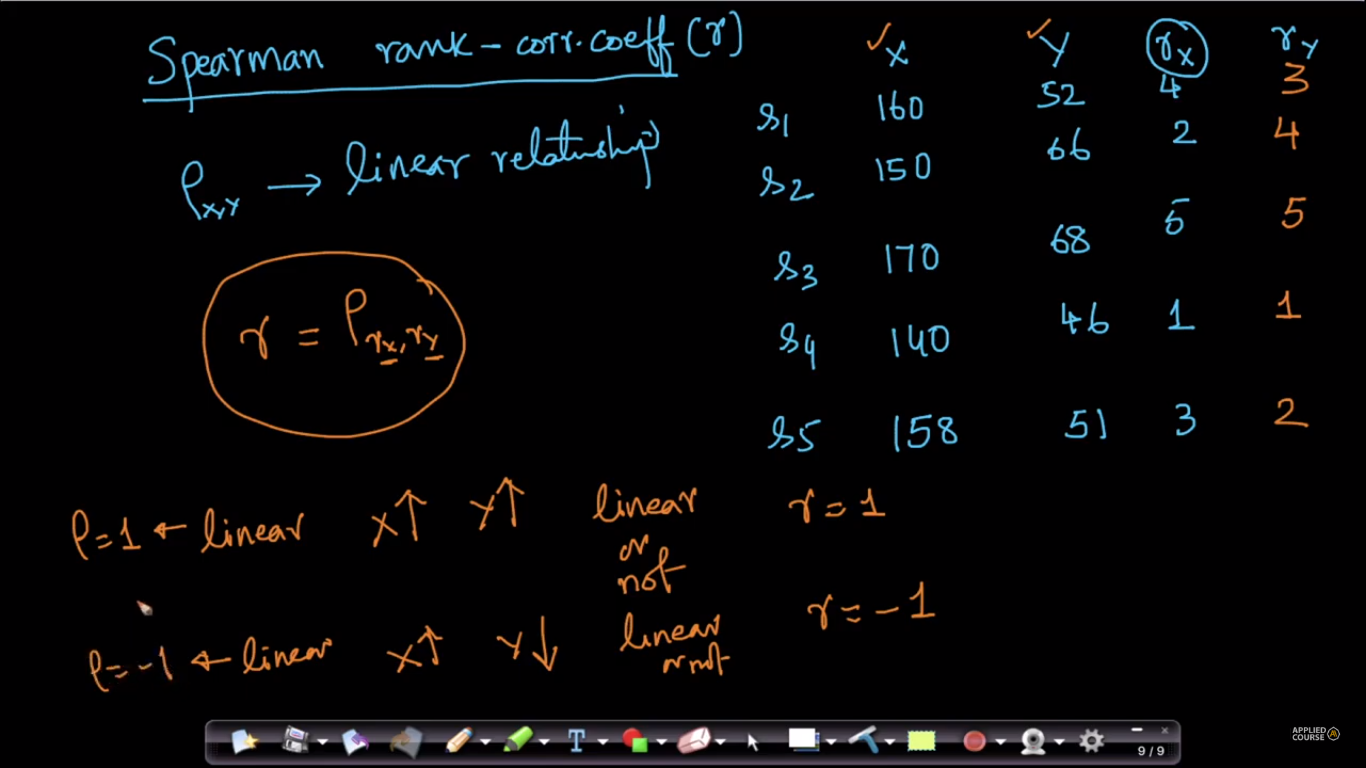


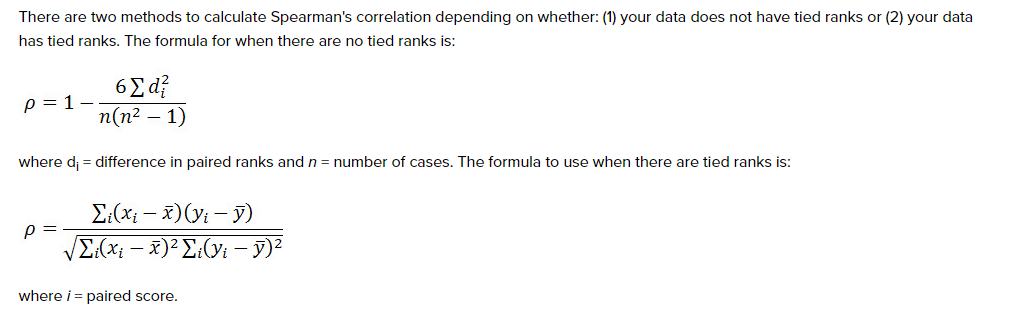
In order to find Spearman correlation we need to give rank to each observation of our RV.

* Assign number 1 to n (the number of data points) corresponding to the variable values in the order highest to lowest.
* In the case of two or more values being identical, assign to them the arithmetic mean of the ranks that they would have otherwise occupied.

For example, Selling Price values given: 28.2, 32.8, 19.4, 22.5, 20.0, 22.5 The corresponding ranks are: 2, 1, 5, 3.5, 4, 3.5 The highest value 32.8 is given rank 1, 28.2 is given rank 2,…. Two values are identical (22.5) and in this case, the arithmetic means of ranks that they would have otherwise occupied (3+4/2) has to be taken.

Then we plot the corresponding x and y ranks.





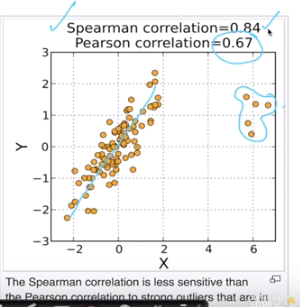
**Advantage of Pearson correlation over Spearman correlation:**

Below given a fig with some outliers, and also mentioned the spearman correlation = 0.84 and pearson correlation = 0.67, that means Pearson correlation works more better with outliers.

But Why?

Because we are not finding the relationship with respect to actual observations value, and we are using ranks provided to them in Spearman correlation, and hence it performs better even if outliers are present.

graph either be linear or non linear; SRCC =1 , but it doesnot implies that it is not effected by outliers, it is too but effect is less significant as compared to PCC. So, If the data has outliers, a few values are far away from others, use Spearman correlation coefficient



**How to use correlation for feature selection:**

<https://towardsdatascience.com/feature-selection-correlation-and-p-value-da8921bfb3cf>