

Anscombe's Quartet

Anscombe's Quartet are 4 datasets that were constructed in 1973 by the statistician Francis Anscombe. We are going to spend a bit of time exploring them.

Set1		Set 2		Set 3		Set 4	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.2	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

For each data set compute the following summary statistics:

- The mean of X
- The Mean of Y
- The variance of X
- The Variance of Y
- The correlation coefficient between X and Y
- The equation of the linear regression line between X and Y

What do you notice with these summary statistics?

What does this lead you to believe about the underlying distributions of the data?

Now plot each data set on a graph of the x/y coordinate plane – what do you see? Is this what you were expecting?