

02. Motivation for Data Visualization

Summary Statistics vs. Visualizations

Summary statistics like the mean and standard deviation can be great for attempting to quickly understand aspects of a dataset, but they can also be misleading if you make too many assumptions about how the data distribution looks.

Anscombe's Quartet Example

Consider we have the following four datasets of x, y pairs. You can download the data using the button below. A link to a Google Sheet with the data is also available [here](#).

DOWNLOAD DATA

I		II		III		IV	
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	7.24	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

QUIZ QUESTION::

Use the data above to match an answer to each of the following questions. (Assume rounding to 2 digits)

ANSWER CHOICES:

They are the same.

They are the same.

They are the same.

They are different.

They are the same.

They are different.

They are different.

They are different.

Question

Answer

What is true for the means associated with any of the **X** columns?

The same

What is true for the means associated with any of the **Y** columns? [The same \(.2f\)](#)

What is true for the standard deviation associated with any of the **X** columns? [The same](#)

What is true for the standard deviation associated with any of the **Y** columns? [The same \(.2f\)](#)

[Next Concept](#)