Overview: Descriptive Statistics

1. **What** are descriptive statistics?

2. Descriptive vs **inferential** statistics

3. **Why** the descriptive matter

4. The "**Big 7**" descriptive statistics

Measures of central tendency

To make this a little more tangible, let's look at a sample dataset, along with the corresponding mean, median and mode. This dataset reflects the service ratings (on a scale of 1 – 10) from 15 customers.

ID	Service Rating (1-10)
Customer 1	3
Customer 2	5
Customer 3	6
Customer 4	10
Customer 5	5
Customer 6	5
Customer 7	9
Customer 8	6
Customer 9	8
Customer 10	4
Customer 11	6
Customer 12	7
Customer 13	4
Customer 14	2
Customer 15	7

Measures of central tendency	
Mean (Average)	5.80
Median	6
Mode	5

Measures of dispersion	
Range	8
Variance	4.74
Std. Deviation	2.18

GRADCOACH

As you can see, the **mean of 5.8** is the average rating across all 15 customers. Meanwhile, **6 is the median**. In other words, if you were to list all the responses in order from low to high, Customer 8 would be in the middle (with their service rating being 6). Lastly, the number **5 is the most frequent rating** (appearing 3 times), making it the mode.

To take this a step further, let's look at the **frequency distribution of the responses**. In other words, let's count how many times each rating was received, and then plot these counts onto a bar chart.
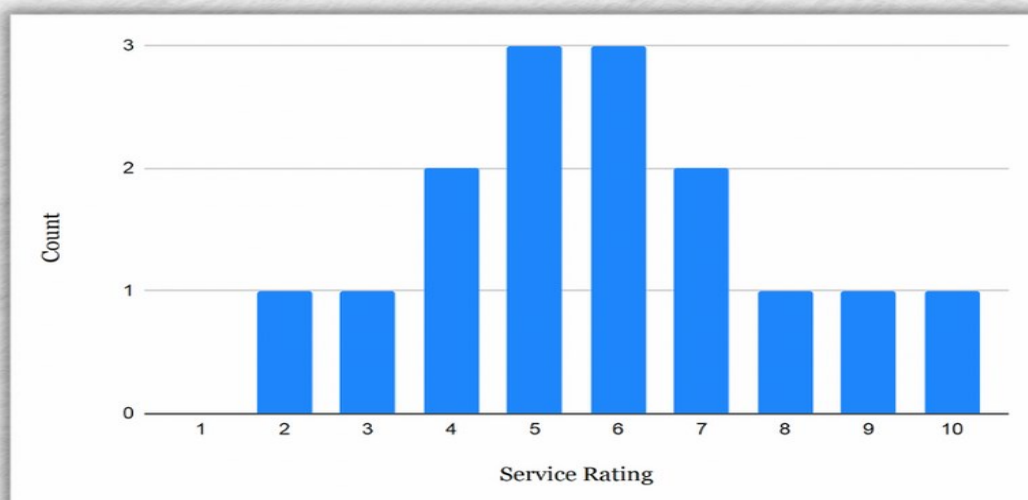
As you can see, the responses **tend to cluster toward the centre of the chart**, creating something of a bell-shaped curve. In statistical terms, this is called a **normal distribution**.

As you delve into quantitative data analysis, you'll find that **normal distributions are very common**, but they're certainly not the only type of distribution. In some cases, the **data can lean toward the left or the right** of the chart (i.e., toward the low end or high end). This lean is reflected by a measure called **skewness**, and it's important to pay attention to this when you're analysing your data, as this will have an impact on what types of inferential statistics you can use on your dataset.
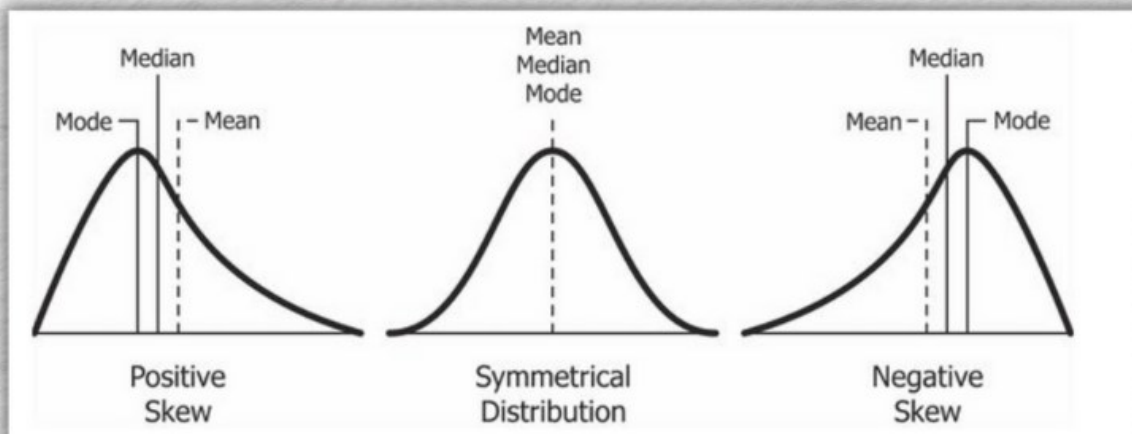
Measures of dispersion

Again, let's look at our sample dataset to make this all a little more tangible.
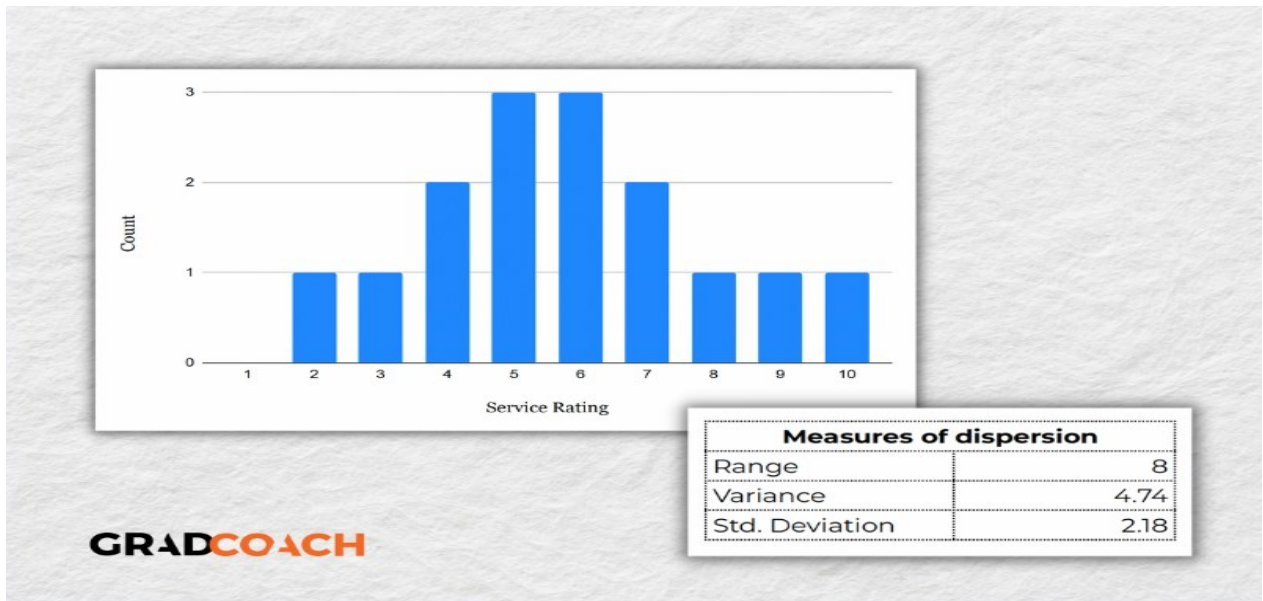
As you can see, the **range of 8** reflects the difference between the highest rating (10) and the lowest rating (2). The **standard deviation of 2.18** tells us that on average, results within the dataset are 2.18 away from the mean (of 5.8), reflecting a **relatively dispersed set of data**.

For the sake of comparison, let's look at another much more **tightly grouped (less dispersed) dataset.**
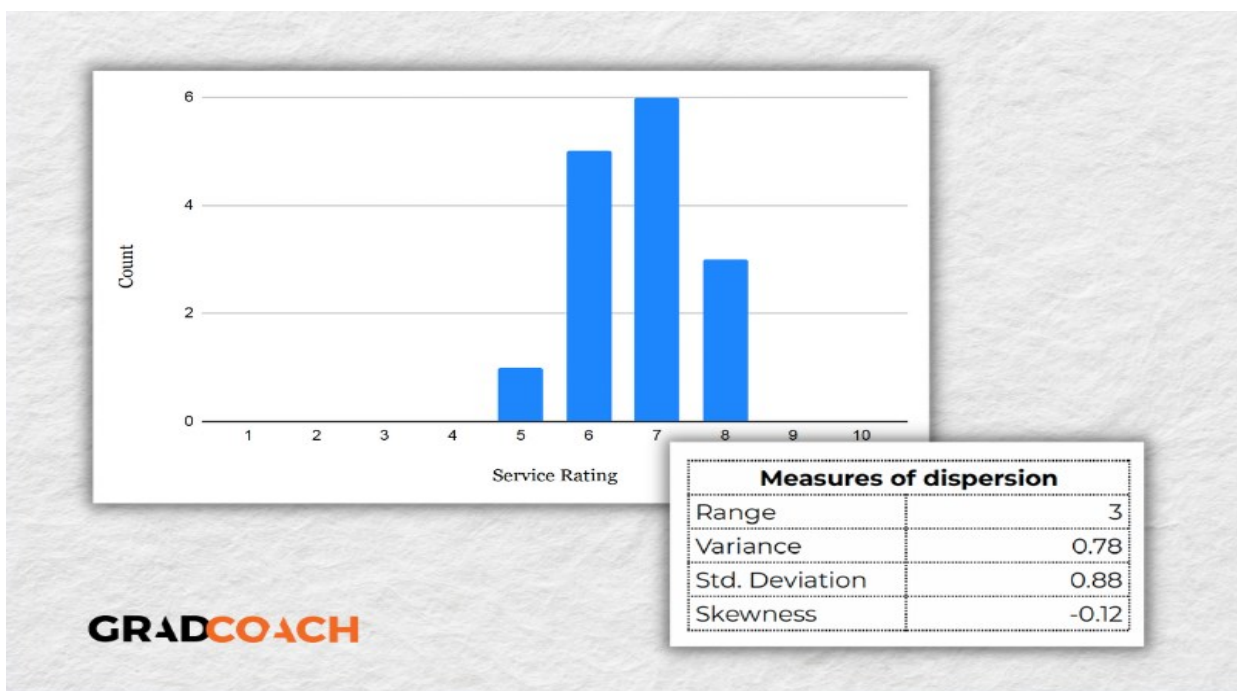
As you can see, all the ratings lay between 5 and 8 in this dataset, resulting in a **much smaller range, variance and standard deviation**. You might also notice that the data are **clustered toward the right side of the graph** – in other words, the data are skewed. If we calculate the skewness for this dataset, we get a result of -0.12, confirming this right lean.

In summary, range, variance and standard deviation all **provide an indication of how dispersed the data are**. These measures are important because they **help you interpret the measures of central tendency within context**. In other words, if your measures of dispersion are all fairly high numbers, you need to **interpret your measures of central tendency with some caution**, as the results are not particularly centred. Conversely, if the data are all tightly grouped around the mean (i.e., low dispersion), the mean becomes a much more "meaningful" statistic).

5. [Key **takeaways**](#)

We've covered quite a bit of ground in this post. Here are the key takeaways:

- Descriptive statistics, although relatively simple, are a **critically important** part of any quantitative data analysis.

- **Measures of central tendency** include the mean (average), median and mode.

- **Skewness** indicates whether a dataset leans to one side or another

- **Measures of dispersion** include the range, variance and standard deviation