Cheatsheet - Statistics

Fabio Lama – fabio.lama@pm.me

1. The Arithmetic Mean

The **arithmetic mean**, or **mean**, is the first average and is found in a set of values by adding up all the values and dividing the result by the total number of values in the set.

$$\text{mean} = \frac{\text{sum of the values}}{\text{total number of values}}$$

For example, the sum of all marks is $5+8+8+6=27$, where the number of marks is 4, hence:

$$\frac{27}{4} = 6.75$$

In more advanced math, we say we have n values and call those $x_1, x_2, \ldots, x_n$ where the *mean* is given as $\bar{x}$ ("x bar"). We sum up all those values and divide it by *n*:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

1.1. Frequencies

When the data is presented in the form of a frequency distribution the mean is found by first multiplying each data value by its frequency and the mean is found by dividing this sum by the sum of all the frequencies

The frequencies are given as:

$$\begin{gathered} f_1 \to x_1 \\ f_2 \to x_2 \\ f_3 \to x_3 \\ \ldots \\ f_n \to x_n \end{gathered}$$

For example, a frequency distribution would be a table that shows how many students (f_n) got which mark (x_n):

$$\begin{gathered} 0 \to 0 \quad (0 \text{ students got mark } 0) \\ \ldots \\ 15 \to 7 \quad (15 \text{ students got mark } 7) \\ 12 \to 8 \quad (12 \text{ students got mark } 8) \\ \ldots \end{gathered}$$

And is calculated as:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \times x_i}{\sum_{i=1}^{n} f_i}$$

2. The Median

The **median** is the second average and is found in a set a set of values by listing the numbers in ascending order and the selecting the value that lies halfway along the list. If there are two numbers halfway, implying the size of the set is even, then the two numbers are averages (i.e. the mean).

For example, in the set $\{1, 4, 5, 7, 9\}$ the value haflway is 5, making that value the **median** of the set. Additionally, in the set $\{3, 4, 5, 6, 7, 7\}$ there are two values halfway, 5 and 6, and we can calculate the mean in order to receive the median:

$$\frac{5+6}{2} = \frac{11}{2} = 5.5$$

3. The mode

The **mode** is the third average and indicates in a set of values what value occurs the most often. For example, in the set of values $\{1, 2, 3, 4, 4, 4, 5, 5\}$ the value 4 appears the most, making it the mode. Sets that have two modes are called **bimodal**.

4. Variance and Standard Deviation

The means of sets can be the same even though the values are widely spread, for example:

$$\begin{gathered} 4+7+10 = 7 \\ 7+7+7 = 7 \end{gathered}$$

If this spread/information should be considered, then the **variance** and **standard deviation** is used.

4.1. Variance

The variance shows the spread between the values.

$$\text{variance } = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

Respectively:

$$\text{variance } = \frac{(4-7)^2 + (7-7)^2 + (10-7)^2}{3} = \frac{18}{3} = 6$$

and

$$\text{variance } = \frac{(7-7)^2 + (7-7)^2 + (7-7)^2}{3} = 0$$

Even though the mean between the two sets are the same, the variance shows us that the spread differs quite a lot between both sets.

In case of frequency distribution:

$$\text{variance } = \frac{\sum_{i=1}^{n} f_i(x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}$$

4.2. Standard Deviation

The standard deviation shows the deviation from the mean, indicating which values are "standard", and which values are "non-standard" respectively vary a lot from most values.

$$\text{standard deviation } = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

The standard deviation is just the square root of the variance. From the examples in the variance section, we can conclude:

$$\text{standard deviation } = \sqrt{6} \approx 2.449$$

and

$$\text{standard deviation } = \sqrt{0} \ = 0$$

Any values above ($\bar{x} + \text{standard deviation}$) and below ($\bar{x} - \text{standard deviation}$) the mean are considered non-standard.

Last updated 2022-09-05 16:20:21 UTC