Cheatsheet - Statistics

Fabio Lama – fabio.lama@pm.me

1. The Arithmetic Mean

The **arithmetic mean**, or **mean**, is the set of values found 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:

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

In more advanced math, we say we have n values and call those x_1, x_2, ..., 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:

$$f_1 \rightarrow x_1$$
$$f_2 \rightarrow x_2$$
$$f_3 \rightarrow x_3$$
$$\dots$$
$$f_n \rightarrow x_n$$

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

$$0 \rightarrow 0 \quad (0 \text{ students got mark } 0)$$
$$\dots$$
$$15 \rightarrow 7 \quad (15 \text{ students got mark } 7)$$
$$12 \rightarrow 8 \quad (12 \text{ students got mark } 8)$$
$$\dots$$

And is calculated as:

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

2. The mode

A third average is the **mode**, which 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**.

3. Variance and Standard Deviation

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

$$4 + 7 + 10 = 7$$
$$7 + 7 + 7 = 7$$

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

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

In case of frequency distribution:

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

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

TODO...

Last updated 2022-08-20 19:41:01 UTC