Standard deviation of the sample means

Suppose you throw a dice. What number will you get? You will get a score of 1, 2, 3, 4, 5 and 6, all with an equal probability. If you denote this number by x, what is its expectation value $\langle x \rangle$?

$$\langle x \rangle = \frac{1+2+3+4+5+6}{6} = 3.5 \tag{1}$$

What is its variance of score?

$$\mathrm{Var}(x) = \frac{(-2.5)^2 + (-1.5)^2 + (-0.5)^2 + 0.5^2 + 1.5^2 + 2.5^2}{6} \approx 2.917 \tag{2}$$

And its standard deviation?

$$\sigma_x = \sqrt{\mathrm{Var}(x)} \approx 1.71 \tag{3}$$

This value seems reasonable. If you throw the dice once, you get 1 from 6. When you get 1 or 6, the deviation from the average is 2.5. However, they are the cases of the maximum deviation, so the standard deviation must be smaller than 2.5. Indeed 1.71 is smaller than 2.5.

Now suppose you throw the dice twice, and your score is the sum of the two numbers you get. There are $36 (= 6 \times 6)$ cases, as the first number can be 1 to 6 and the second number can be 1 to 6. See Table 1. The red number denotes the result of the first try, and the blue number denotes the result of the second try. The black number is your score, which is given by the sum of the red number and the blue number. What is the expectation value of your score? There are two ways to calculate this, a stupid one and a clever one.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 1: Red numbers denote the first try and blue numbers denote the second try. Black numbers denote the scores.

Let's try the stupid one first. There are 36 cases, each with an equal probability, so you add all the black numbers in Table 1 and divide by 36. Then, you will get 7.

Let's try now the clever one. The expectation value of the first throw will be 3.5, and the expectation value of the second throw will be 3.5. As your score is the sum of the first throw and the second throw, the answer is 7(=3.5+3.5).

What are the variance and the standard deviation of the score? There are two ways to calculate this, a stupid one and a clever one.

Let's try the stupid one first. As the average is 7, you subtract 7 from each black entry in Table 1, square them, sum all of them, and divide by 36. It's a complicated calculation, so I don't recommend you to try this method.

Let's try now the clever one. Note that the first throw and the second throw are independent; no matter what value you got for the first throw, the probability that you will get 1, or 2, or 3, or 4, or 5, or 6 for the second throw will be all the same. If you remember what we have learned in our earlier article, you will see that the variance of your score, the black numbers will be the sum of variance of the red numbers and variance of the blue numbers. Thus, we get

$$\mathrm{Var}(\mathrm{score}) = \mathrm{Var}(\text{first try}) + \mathrm{Var}(\text{second try}) \approx 2.917 + 2.917 \approx 5.833 \tag{4}$$

The standard deviation is given by

$$\sigma_{\mathrm{score}} = \sqrt{5.833\cdots} = 2.415\cdots \tag{5}$$

So, the expectation value of your score is 7 with the standard deviation of 2.415. Can we express this standard deviation in terms of (3)? Yes. As (2) is σ_x^2, (4) is given by

$$\mathrm{Var}(\mathrm{score}) = \sigma_x^2 + \sigma_x^2 = 2\sigma_x^2 \tag{6}$$

and (5) is given by

$$\sigma_{\mathrm{score}} = \sqrt{2\sigma_x^2} = \sqrt{2}\sigma_x \tag{7}$$

What if you throw the dice 20 times, and the score is the sum of these 20 numbers? **Problem 1.** Check that the expectation value of the score you get is 70 with the standard deviation of $7.637\cdots (= \sqrt{20}\sigma_x)$.

In general, if you throw the dice N times, and the score is the sum of these N numbers, the score you get is $3.5N$ with the standard deviation of $\sqrt{N}\sigma_x$.

Suppose now, if you throw the dice N times, and the score is *not* the sum of these N numbers, but the *average* of these N numbers. What is the expectation value and the standard deviation of the score? The expectation value is $3.5 = 3.5N/N$, and the standard deviation is

$$\frac{\sqrt{N}\sigma_x}{N} = \frac{\sigma_x}{\sqrt{N}} \tag{8}$$

Notice that the bigger N, the smaller the standard deviation. What does this mean?

Suppose you throw the dice 100 times. Then, you expect that the average of the 100 numbers you obtained must be much closer to 3.5 than the average from throwing 5 times or 10 times. Let me explain what I mean. If you throw small number of times, you can get a big average (or a small average), because you have gotten, say, too many 6s (or 1s), by chance. However, if you throw 100 numbers, too many 6s (or too many 1s) are averaged out, as the chance that you get too many 6s (or too many 1s) becomes smaller. A careful reader might have noticed this already. In case of a single throw, the chance of getting a 6 is 1/6. However, in case of two throws, the chance of getting two 6s is 1/36, as you can see from Table 1. 1/36 is much smaller than 1/6.

This can be said mathematically more precisely. According to (8), the distribution of the average with 100 throws is the mean of 3.5 with the standard deviation 0.17, whereas the same quantities for the case of 5 times and 10 times are mean with both 3.5 and the standard deviation of 0.76 and 0.54, respectively. In other words, while the average of 100 throws is most likely to deviate from the mean 3.5 by the amount comparable to 0.17, if you average only 5 throws (or 10 throws), the average is most likely to deviate from the mean 3.5 by the amount comparable to 0.76 (or 0.54).

Problem 2. Let's say you throw a coin 10,000 times. The score you get is the average number of head per throw. For example, if you got 5,050 heads, your score is 0.505. Then, what is the expectation value and the standard deviation of your score?

So far, we threw a dice or a coin, but the result so far can be generalized to choosing a sample from a big data. Let's say that the average height of citizens of A is 165 cm with the standard deviation of 15 cm. If you randomly choose 100 citizens from A, and calculate the average, then the distribution of so-obtained average will be the mean of 165 cm with the standard deviation of 1.5 cm ($= 15/\sqrt{100}$). There can be many tall citizens or short citizens, but if you choose as many as 100 citizens, they are averaged out.

What we have learned in this article can have a real application. In 1345, English law set that the weight of all coins must weigh within the deviation of 0.7% of the target weight. [1] Otherwise, the coin manufacturers can pocket the silver that is supposed to be used to make the coins by making light coins. However, when this law was enforced, it was very ineffective, because the coin examiners thought that the deviation of 0.7% of weight of each coin would mean the deviation of 0.7% of the sum of the weight of sample coins to be examined. Let me explain what I mean. Let's say the target weight is 1 gram. Then, the coin examiners thought that the coins are accepted, say, if 100 coins weigh between 99.3 gram and 100.7 gram. At first glance, this may sound correct, but it is not. Let's see why. Suppose that the real distribution of the manufactured coins is the average of 0.995 gram with the standard deviation of 0.01 gram. Certainly, many coins are unacceptable, as some are as light as or lighter than $0.985 (= 0.995 - 0.01)$ gram, which is quite smaller than 0.993 gram, the minimum weight of the coin that can be accepted. (0.985 gram is 1.5 % less than the target weight,

[1] I read about this story from *AIQ: How People and Machines Are Smarter Together* by Nick Polson and James Scott.

while acceptable weight is 0.7 % less than the target weight.) However, the average of 100 coins is 0.995 gram with the standard deviation of 0.001g ($= 0.01/\sqrt{100}$). In other words, the average weight lies somewhere around between 0.994 gram($= 0.995 - 0.001$) and 0.996 gram($= 0.995 + 0.001$). In other words, their total weight lies somewhere around 99.4 gram and 99.6 gram. So, they are most likely to be falsely accepted, as the total mass is most likely to be higher than 99.3 gram, even though many of each individual coin weigh less than 0.993 gram.

Final comment. Consider the case in Problem 2. Let's say that you throw a coin 10,000 times, and got the head 6,000 times. Then, we can immediately conclude that this is not an honest coin with the probability of 1/2 for the head. How about if we got the head 5,500 times? How about 5,100 times? How about 5,050 times? You are not sure what is the boundary between an honest coin and a fake coin. In such cases, statisticians can calculate the probability that the coin is a fake one. For example, in case of 6,000 heads out of 10,000 tosses, this probability is very close to 100 %. For 5,100 heads, 5,100 heads or 5,050 heads these probabilities are smaller and smaller. Notice also that this probability depends not only on score but number of throws. If you get 6,000 heads out of 10,000 tosses, or 3 heads out of 5 tosses, your score is both 0.6. Nevertheless, the probability that the later case is the false coin is much lower than 100 %. In our later article "Gaussian distribution," we will explain how to calculate such probabilities, especially for the case of 10,000 tosses, i.e., when the number of toss is big.

Summary

- If a quantity has a distribution of mean x and the standard deviation σ, the mean of this quantity from a sample of size N has the distribution of mean x and the standard deviation of $\sigma/\sqrt{N}$.