Transforming random variables

Remember previously that we talked about how our measures of central tendency and spread would change if we shifted or scaled our data set.

Shifting the data set by a constant k means adding k to every value in the data set, or subtracting k from every value in the data set. On the other hand, scaling the data set by a constant k means multiplying or dividing every value in the data set by k.

We learned that shifting the data set would shift the mean, median and mode by the same amount as the constant, but that the range and IQR would stay the same. For example, shifting a data set up by k might look like this:

Original data set	Shifted data set
Mean: 6	Mean: $6+k$
Median: 7	Median: $7+k$
Mode: 3	Mode: $3+k$
Range: 10	Range: 10
IQR: 8	IQR: 8

To this list, let's add standard deviation. When we shift the data set up or down by k units, the standard deviation will stay the same. So

Original data set	Shifted data set

Mean: 6

Median: 7

Mode: 3

Range: 10

IQR: 8

Standard deviation: σ

Mean: $6+k$

Median: $7+k$

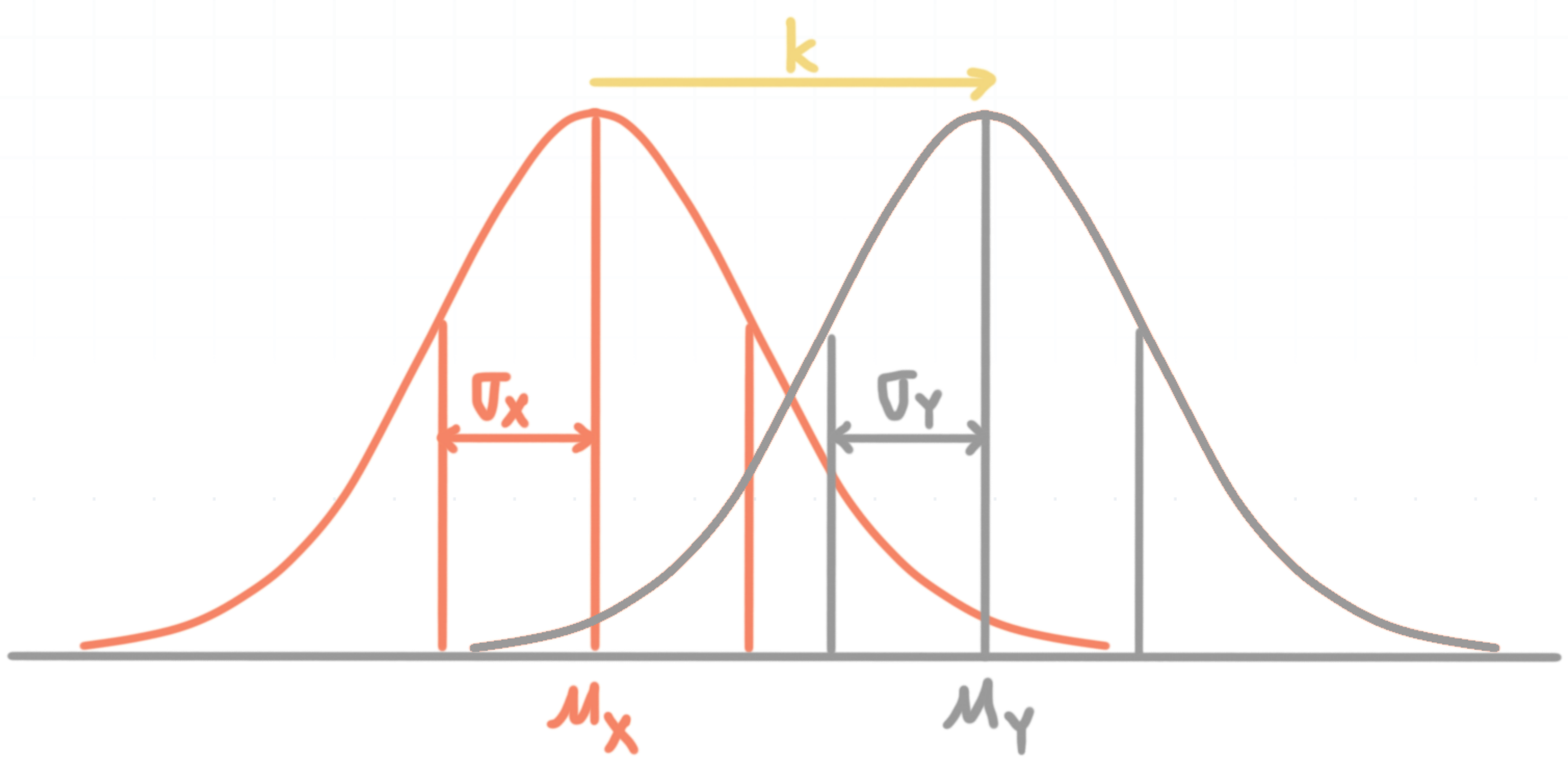
Mode: $3+k$

Range: 10

IQR: 8

Standard deviation: σ

Here's how we'd visually represent shifting in the data.



We also learned that scaling the data set would equally scale the mean, median, mode, range and IQR. In other words, they all scale by the same factor. For example, scaling a data set by multiplying by k might look like this:

Original data set

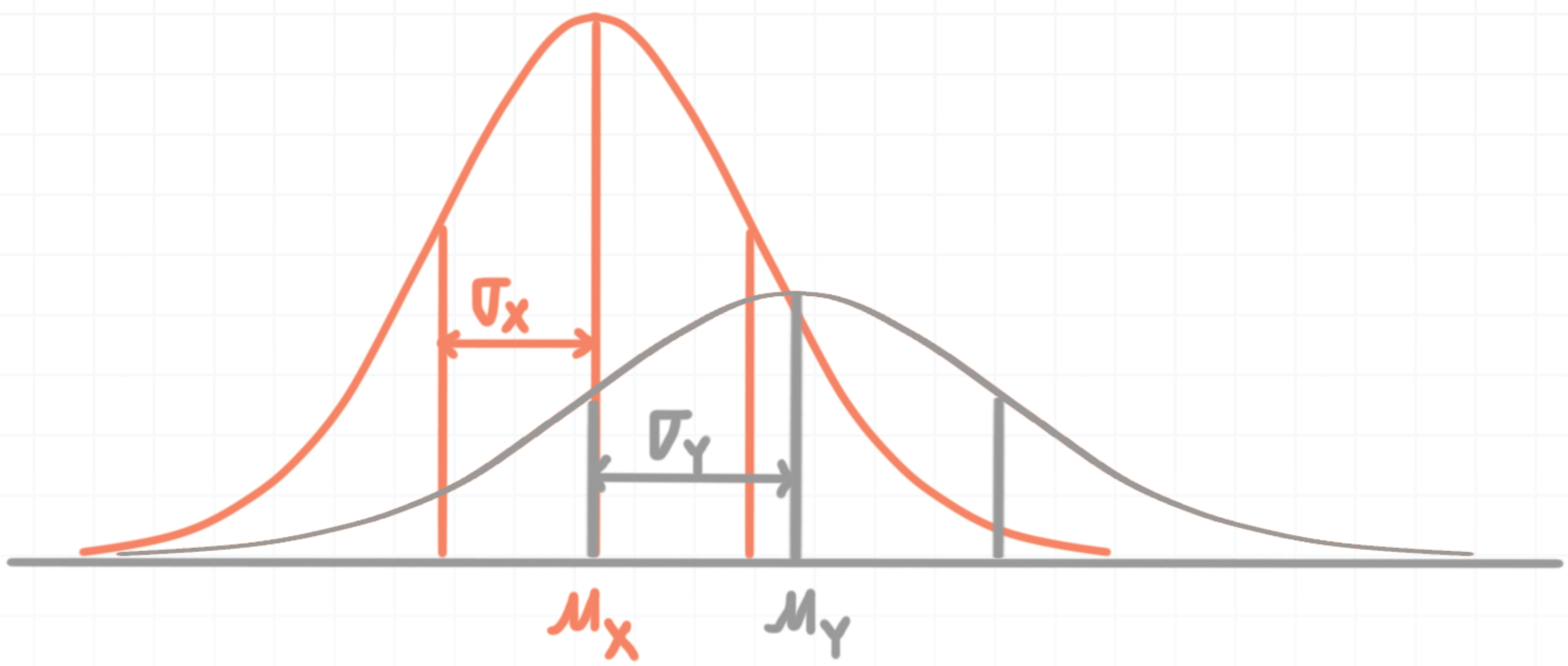
Scaled data set

Mean: 6	Mean: $6k$
Median: 7	Median: $7k$
Mode: 3	Mode: $3k$
Range: 10	Range: $10k$
IQR: 8	IQR: $8k$

But when we scale the data set by k units, the standard deviation will scale by the same value. So

Original data set	Scaled data set
Mean: 6	Mean: $6k$
Median: 7	Median: $7k$
Mode: 3	Mode: $3k$
Range: 10	Range: $10k$
IQR: 8	IQR: $8k$
Standard deviation: σ	**Standard deviation:** σk

Here's how we'd visually represent scaling in the data. The original distribution would become the new scaled version:



Let's do an example.

Example

We're playing a game in which we pay $\$5$ for the chance to shoot a basketball two times. We win $\$3$ for every shot we make. Find the mean and standard deviation of the profit (or loss) we can expect if we play this game, assuming the table represents the probability distribution of X, the number of shots we make.

X	0	1	2
P(X)	0.25	0.49	0.26

The mean, or expected value, is

$$E(X) = \mu_X = 0(0.25) + 1(0.49) + 2(0.26)$$

$$E(X) = \mu_X = 0.49 + 0.52$$

$$E(X) = \mu_X = 1.01$$

Then the variance is

$$\sigma_X^2 = (0 - 1.01)^2(0.25) + (1 - 1.01)^2(0.49) + (2 - 1.01)^2(0.26)$$

$$\sigma_X^2 = (1.0201)(0.25) + (0.0001)(0.49) + (0.9801)(0.26)$$

$$\sigma_X^2 = 0.255025 + 0.000049 + 0.254826$$

$$\sigma_X^2 = 0.5099$$

so the standard deviation is

$$\sigma_X \approx 0.7141$$

From here, we can set up a net gain equation. Every time we play the game, if we make X shots, our expected net gain (or loss) can be given by $N(X) = 3X - 5$, because we get $\$3$ for every shot we make, but we have to pay $\$5$ to play. So the net gain (or loss) for every possible value of X is

$$N(0) = 3(0) - 5 = -5$$

$$N(1) = 3(1) - 5 = -2$$

$$N(2) = 3(2) - 5 = 1$$

and the probability distribution is therefore

N	-5	-2	1
P(N)	0.25	0.49	0.26

Although we could calculate the mean and standard deviation using this new table, let's do it instead by using $N(X) = 3X - 5$ as a transformation of the variable X. In $N(X) = 3X - 5$, X is scaled first, $3X$, and then shifted by -5.

The mean is affected by both scaling and shifting, so the mean of the net gain is

$$\mu_N = 3\mu_X - 5$$

$$\mu_N = 3(1.01) - 5$$

$$\mu_N = 3.03 - 5$$

$$\mu_N = -1.97$$

The standard deviation is affected only by scaling, not shifting, so the standard deviation of the net gain is

$$\sigma_N = 3\sigma_X$$

$$\sigma_N = 3(0.7141)$$

$$\sigma_N = 2.1423$$

Therefore, every time we play the game, we expect to lose $\$1.97$, with a standard deviation of 2.1423.