

Skewed Distributions

3 min

As we're moving through the numeric

Preview: Docs Loading link description

[variables](#)

in our musician dataset, we come across some interesting details when we inspect the income variable.

1. We notice that the shape of the distribution is different than the shape of the age distribution. There are quite a few musicians with higher incomes that are creating a longer tail on the right side.
2. We also notice that the mean indicates that the typical income is \$34,795. This value seems a little high since most of the incomes seem to be between \$15,000 and \$40,000.

What we have learned is that the income distribution is skewed. A **skewed** distribution is asymmetrical with a steep change in frequency on one side and a flatter, trailing change in frequency on the other. Specifically, the income distribution is right-skewed (also called positively-skewed) because the tail is on the right side.

So why does the mean seem wrong? Remember, the mean is the sum of all the values in the dataset divided by the total count. That sum is made very large by all the higher incomes in that right tail. This makes the mean a greater number than we would like it be. When the data are skewed, the mean may not be the best measure of a typical observation.

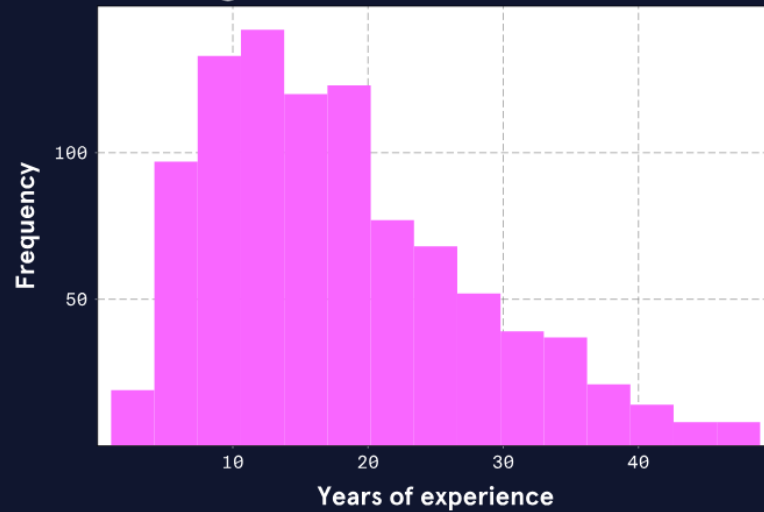
There are a number of ways to deal with this issue. We will handle the problem with the income data by taking some alternative measurements.

Instructions

Another numeric variable from the musician dataset is years of experience working in the field of music. The learning environment shows a card that flips when you hover over it. The plot on the front shows what the experience distribution might look like if it is right-skewed like the income distribution is. The plot on the back shows what the experience distribution might look like if it is left-skewed.

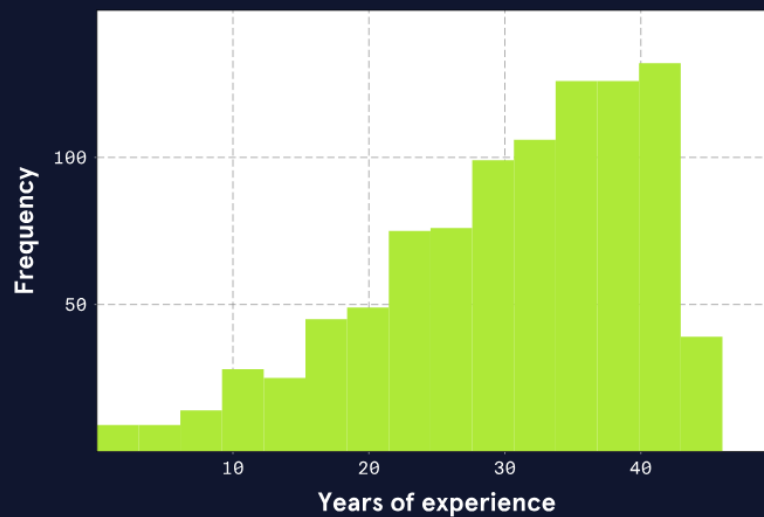
Think about which distribution seems most likely to be true for musicians in Melody Metropolis.

Right-skewed distribution



A right-skewed distribution has a tail on the right. The higher values for experience in the tail will pull the mean out to be **higher** than the median.

Left-skewed distribution



A left-skewed distribution has a tail on the left. The lower values for experience in the tail will pull the mean out to be **lower** than the median.