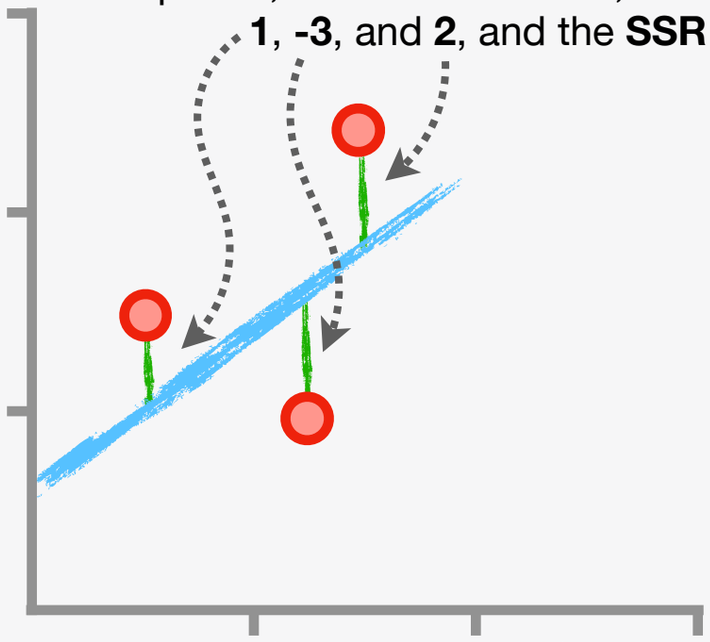
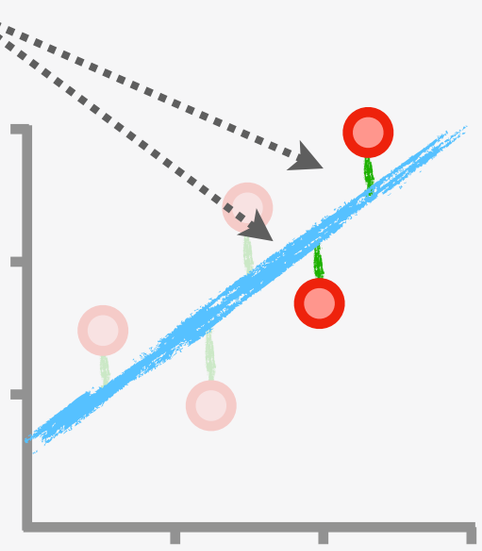
Mean Squared Error (MSE): Main Ideas

1 **The Problem: Sum of the Squared Residuals** (**SSR**), although awesome, is not super easy to interpret because it depends, in part, on how much data you have.

For example, if we start with a simple dataset with **3** points, the **Residuals** are, from left to right, **1**, **-3**, and **2**, and the **SSR = 14**.



Now, if we have a second dataset that includes **2** more data points added to the first one, and the **Residuals** are **-2** and **2**, then the **SSR** increases to **22**.



However, the increase in the **SSR** from **14** to **22** *does not* suggest that the second model, fit to the second, larger dataset, is worse than the first. It only tells us that the model with more data has more **Residuals**.

2 **A Solution:** One way to compare the two models that may be fit to different-sized datasets is to calculate the **Mean Squared Error** (**MSE**), which is simply the average of the **SSR**.

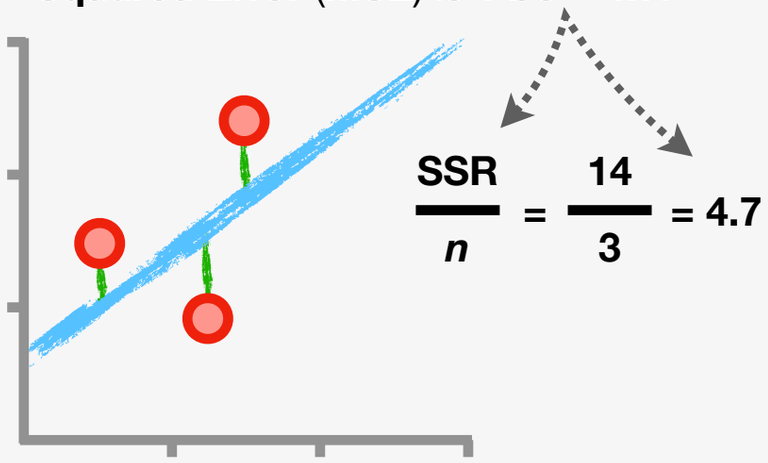
$$\textbf{Mean Squared Error (MSE)} = \frac{\textbf{The Sum of Squared Residuals (SSR)}}{\textbf{Number of Observations, } n} = \sum_{i=1}^{n} \frac{(\text{Observed}_i - \text{Predicted}_i)^2}{n}$$

Mean Squared Error (MSE): Step-by-Step

1 Now let's see the **MSE** in action by calculating it for the two datasets!!!

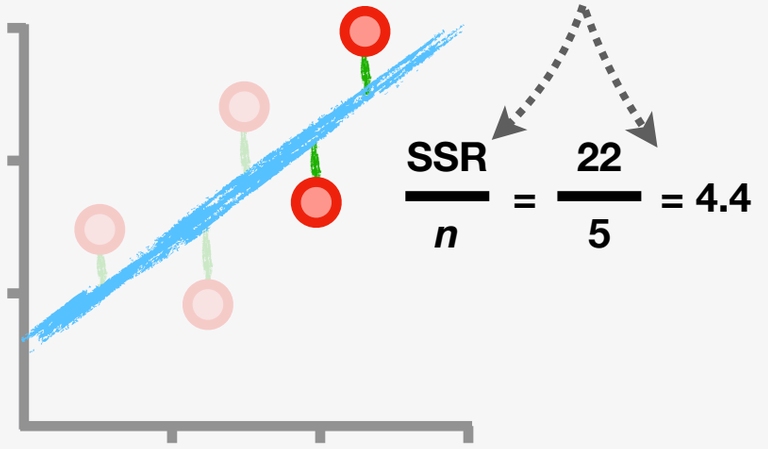
$$\textbf{Mean Squared Error (MSE)} = \frac{\textbf{SSR}}{n} = \sum_{i=1}^{n} \frac{(\text{Observed}_i - \text{Predicted}_i)^2}{n}$$

2 The first dataset has only **3** points and the **SSR = 14**, so the **Mean Squared Error** (**MSE**) is **14/3 = 4.7**.



$$\frac{\textbf{SSR}}{n} = \frac{14}{3} = 4.7$$

The second dataset has **5** points and the **SSR** increases to **22**. In contrast, the **MSE**, **22/5 = 4.4**, is now slightly lower.

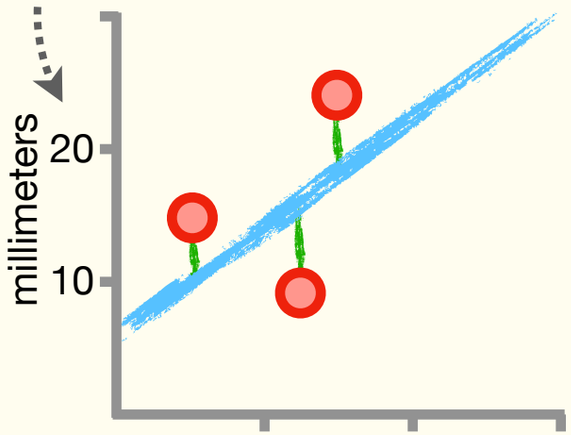


$$\frac{\textbf{SSR}}{n} = \frac{22}{5} = 4.4$$

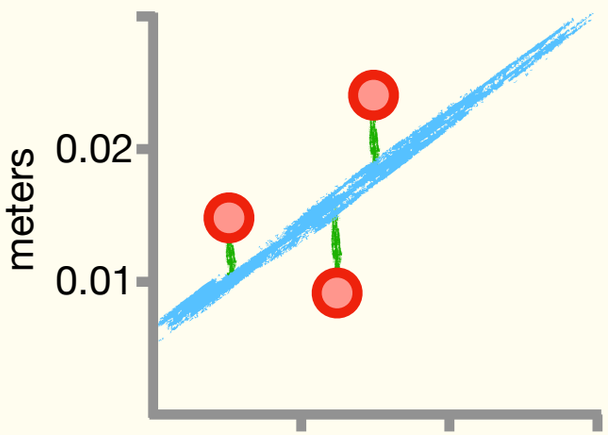
So, unlike the **SSR**, which increases when we add more data to the model, the **MSE** can increase or decrease depending on the average residual, which gives us a better sense of how the model is performing overall.

3 Unfortunately, **MSEs** are still difficult to interpret on their own because the maximum values depend on the scale of the data.

For example, if the y-axis is in *millimeters* and the **Residuals** are **1**, **-3**, and **2**, then the **MSE = 4.7**.



However, if we change the y-axis to *meters*, then the **Residuals** for the exact same data shrink to **0.001**, **-0.003**, and **0.002**, and the **MSE** is now **0.0000047**. It's tiny!



The good news, however, is that both the **SSR** and the **MSE** can be used to calculate something called R^2, which is independent of both the size of the dataset and the scale, so keep reading!