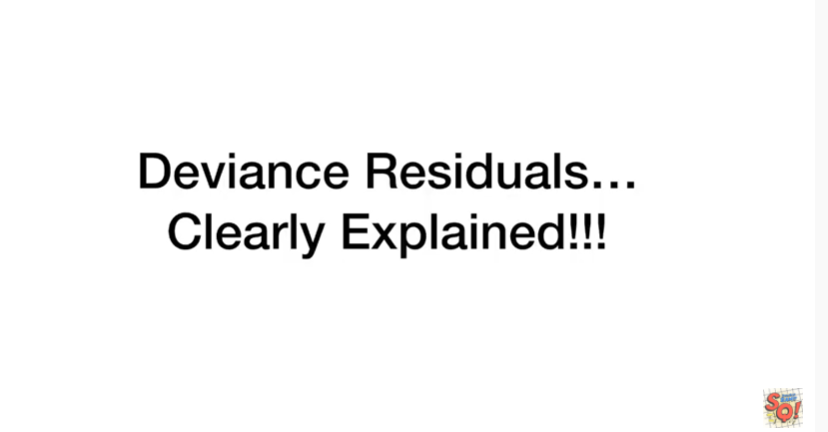
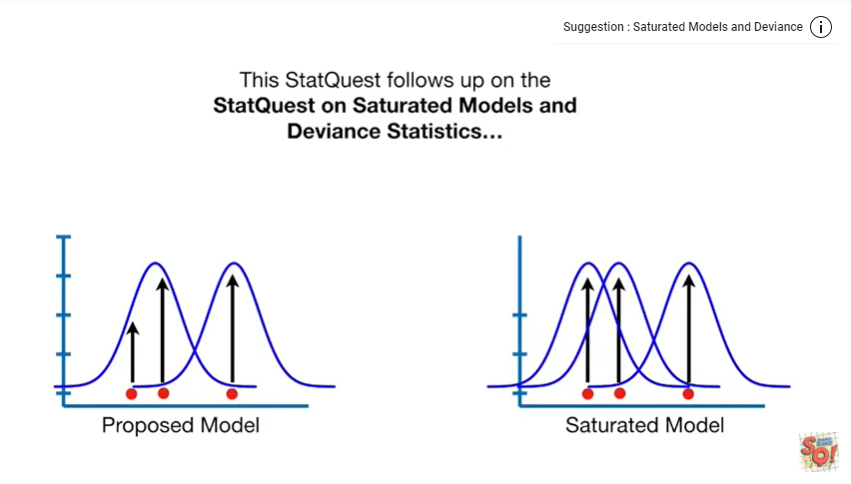
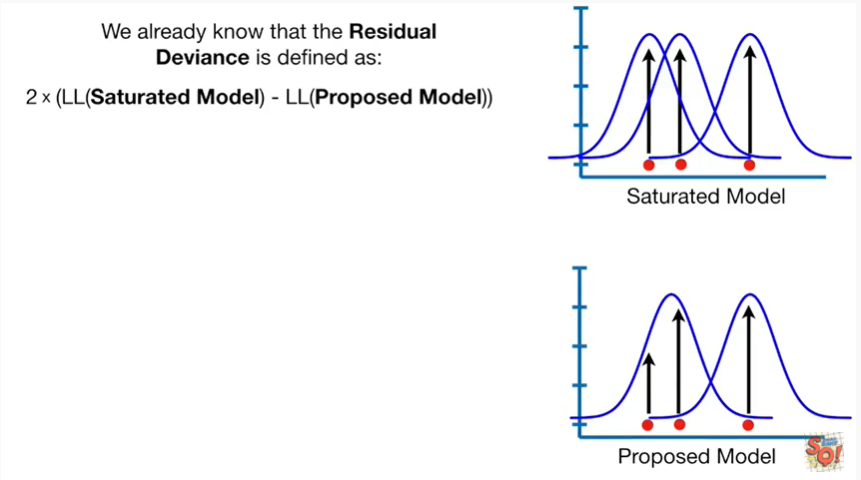
<https://www.youtube.com/watch?v=JC56jS2gVUE&list=PLblh5JKOoLUKxzEP5HA2d-Li7IJkHfXSe&index=6>



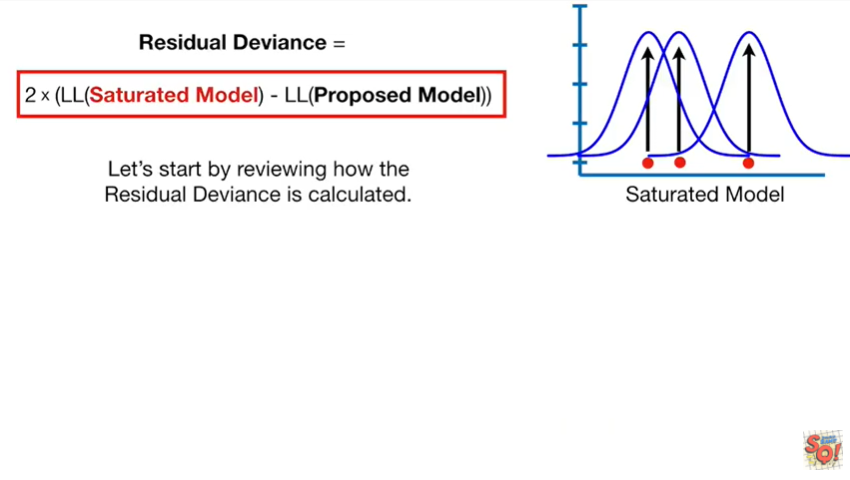
Today we're going to talk about deviance residuals and they're gonna be clearly explained.



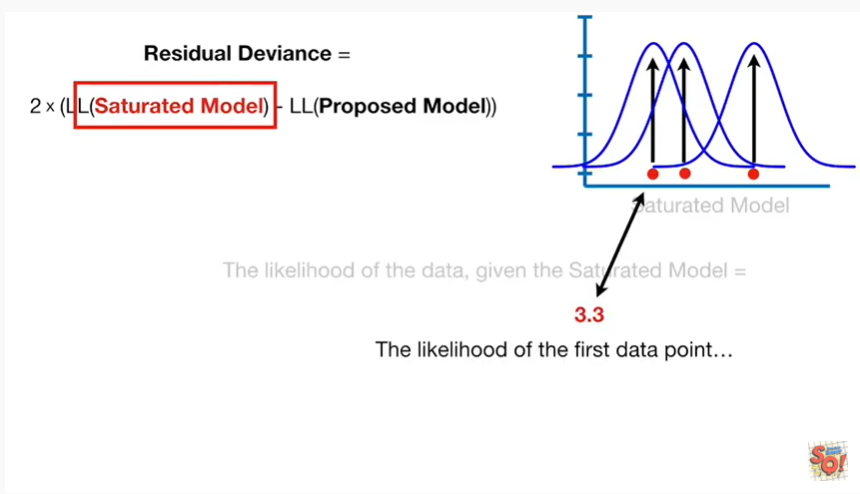
This stat quest follows up on the stat quest on saturated models and deviant statistics so watch that video first if you're not familiar with those topics.



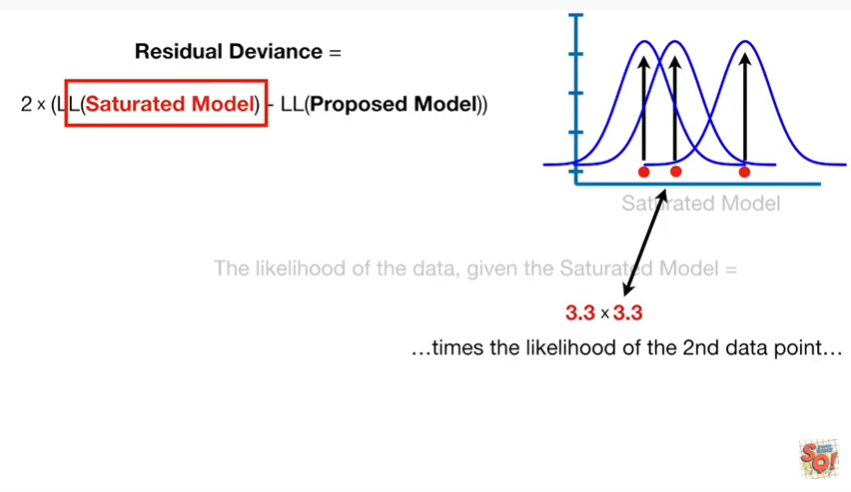
We already know that the residual deviance is defined as two times the difference between the log likelihood of the saturated model and the log likelihood of the proposed model.



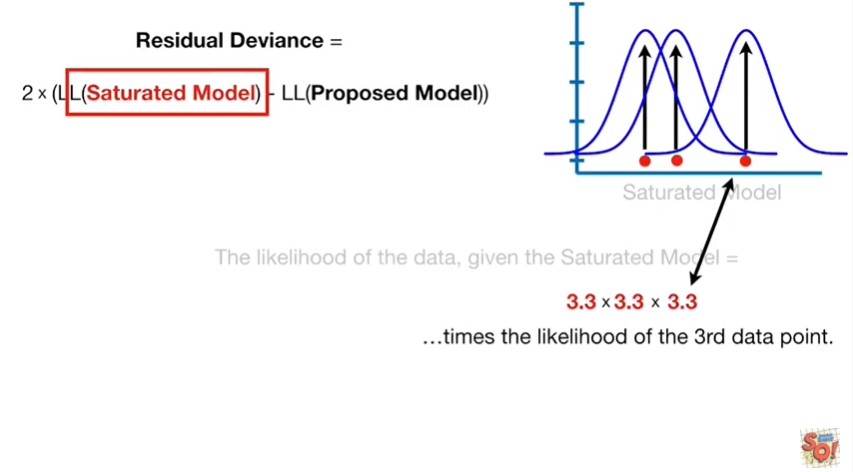
Let's start by reviewing how the residual deviance is calculated the likelihood of the data.



Given the saturated model is equal to the likelihood of the first data point

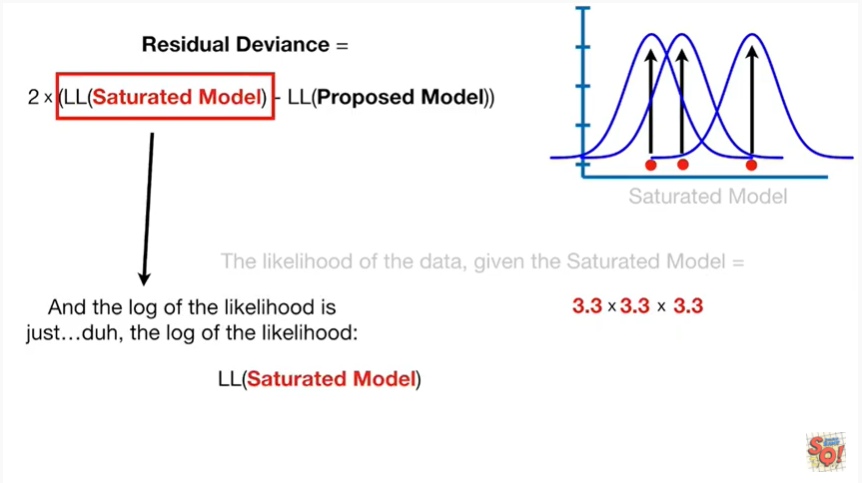


times the likelihood of the second data point

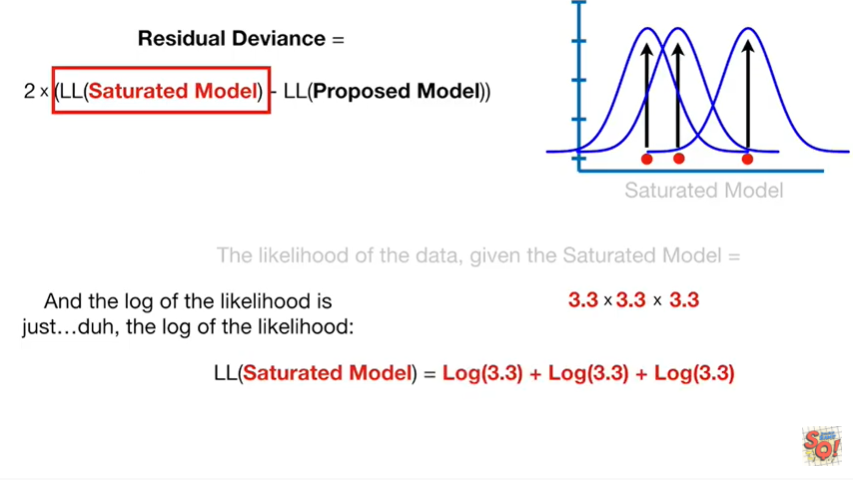


times the likelihood of the third data point.

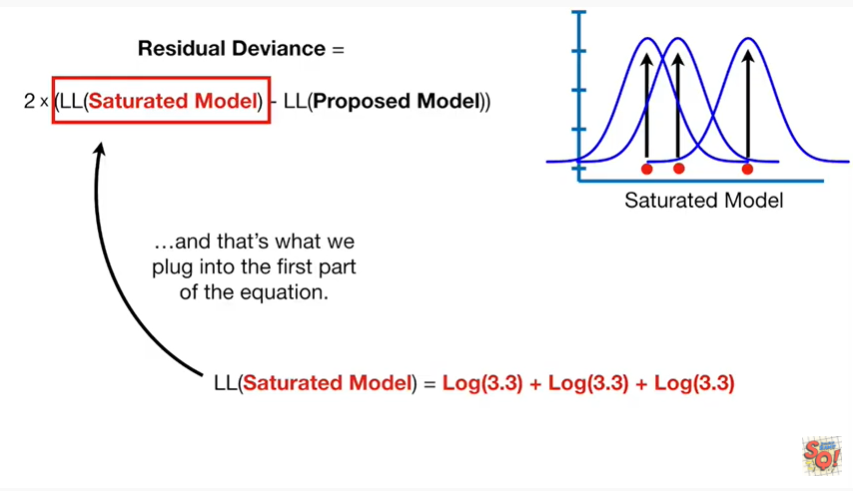
Note : we're just using three data points in this example however if you've got more data you just keep adding them to the multiplication.



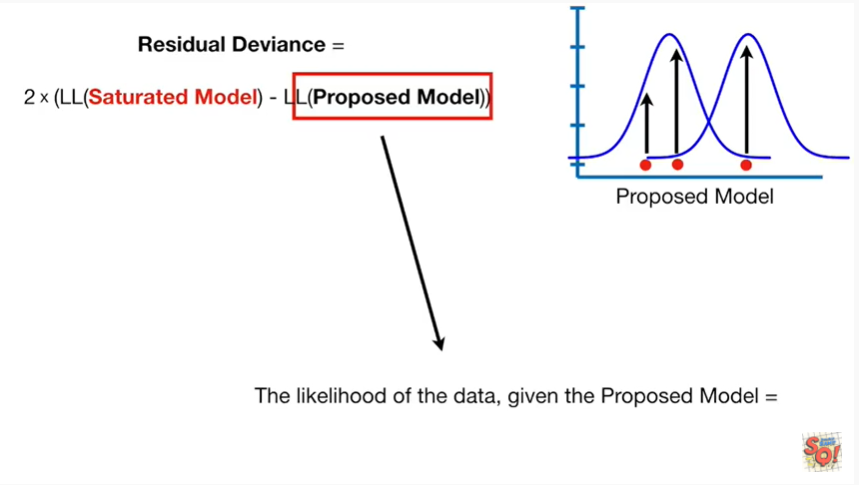
And the log likelihood is just duh



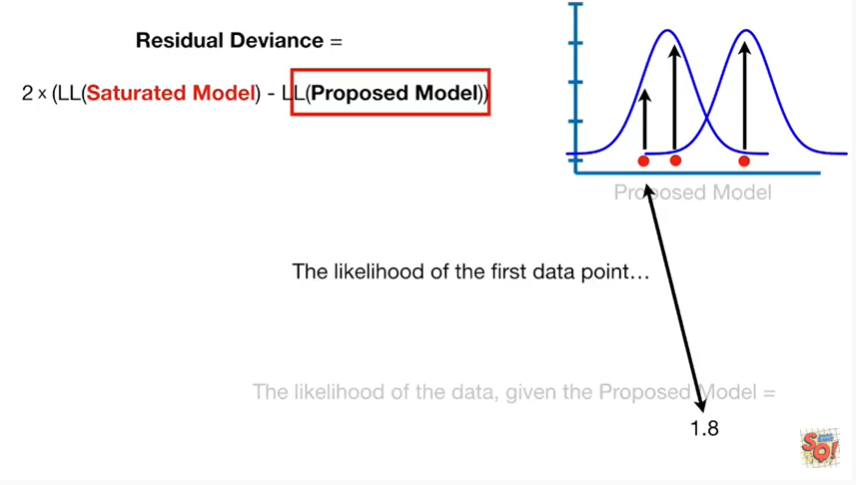
the log of the likelihood



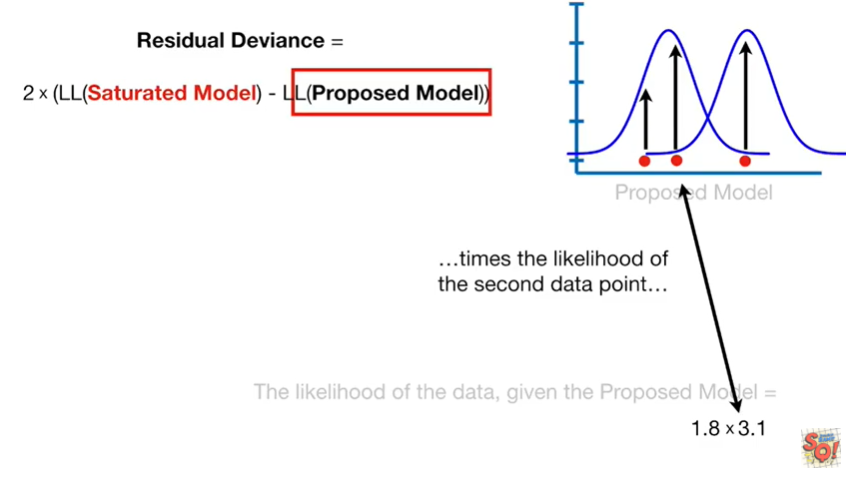
and that's what we plug in to the first part of the equation



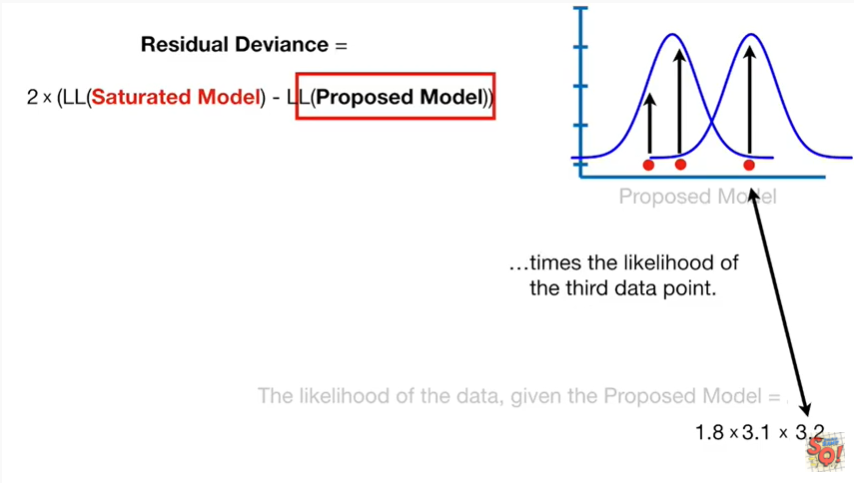
The likelihood of the data given the proposed model equals



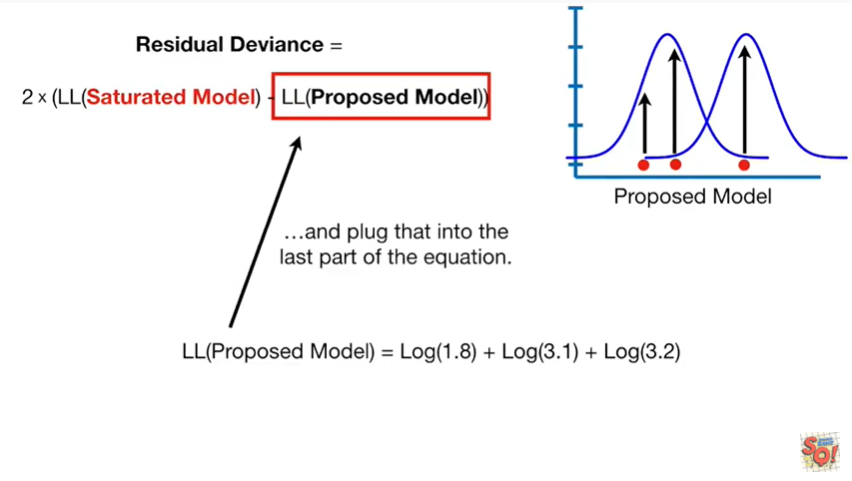
the likelihood of the first data point



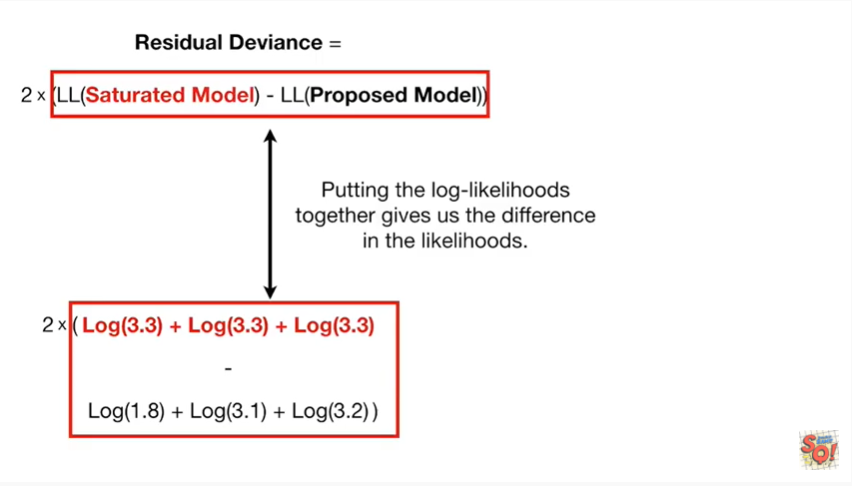
times the likelihood of the second data point



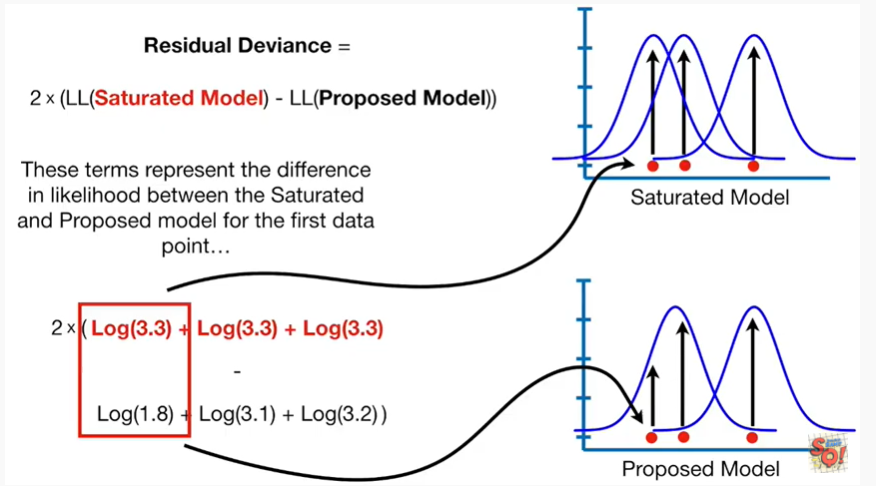
times the likelihood of the third data point.



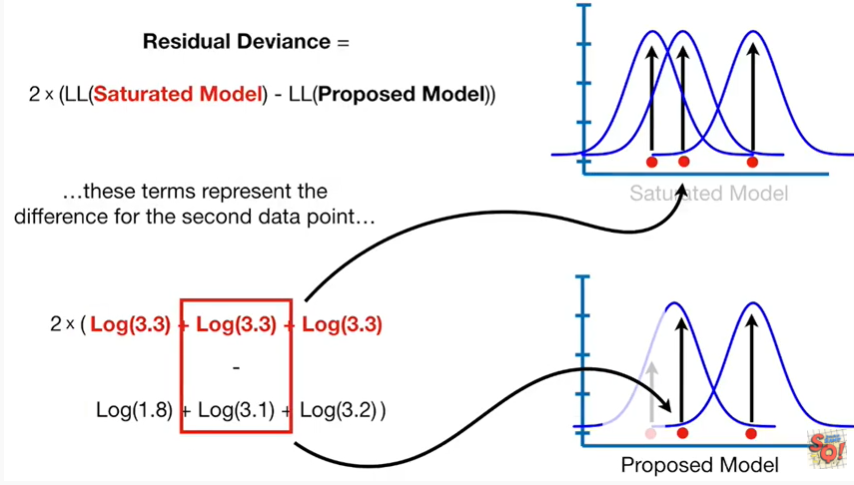
Then we just take the log of the likelihood and plug that into the last part of the equation.



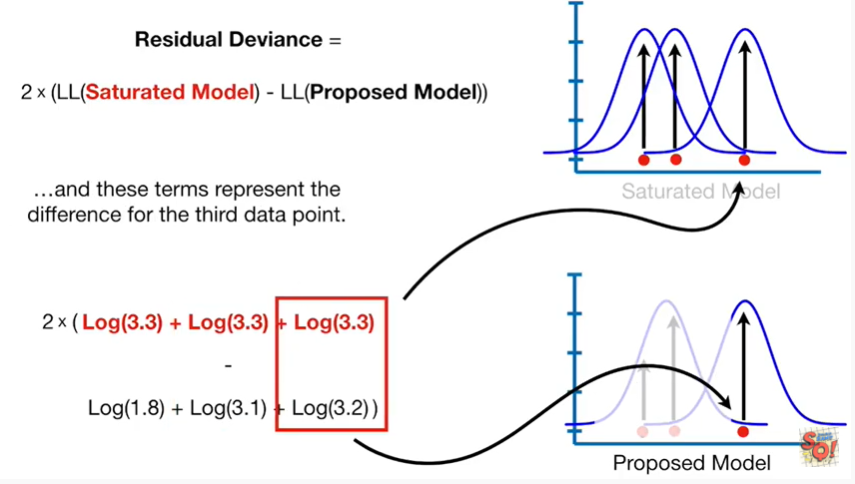
Putting the log likelihoods together gives us the difference in the likelihoods.



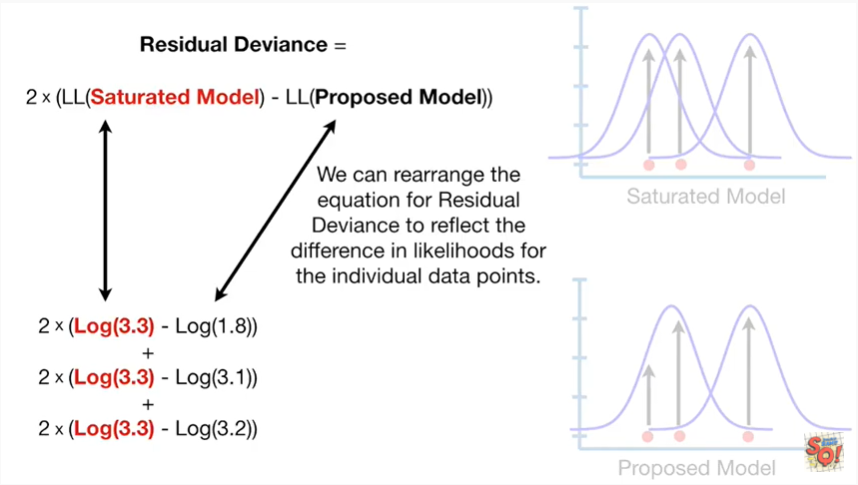
These terms represent the difference in likelihood between the saturated and proposed model for the first data point.



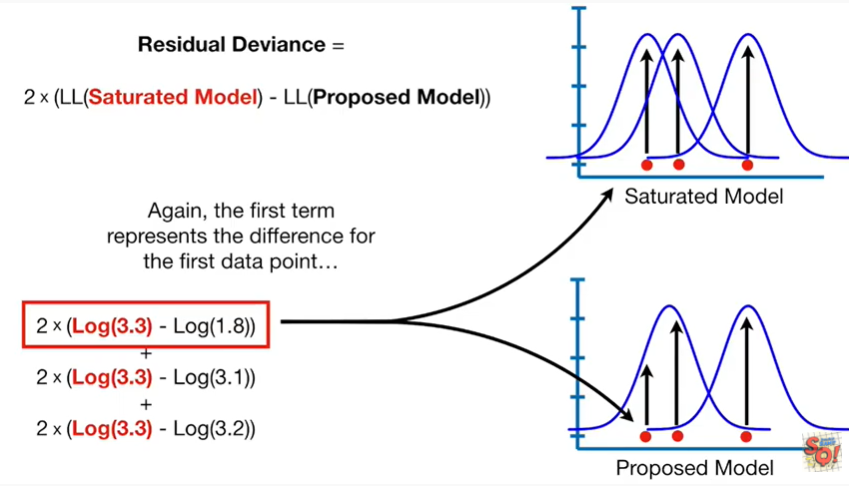
These terms represent the difference for the second data point.



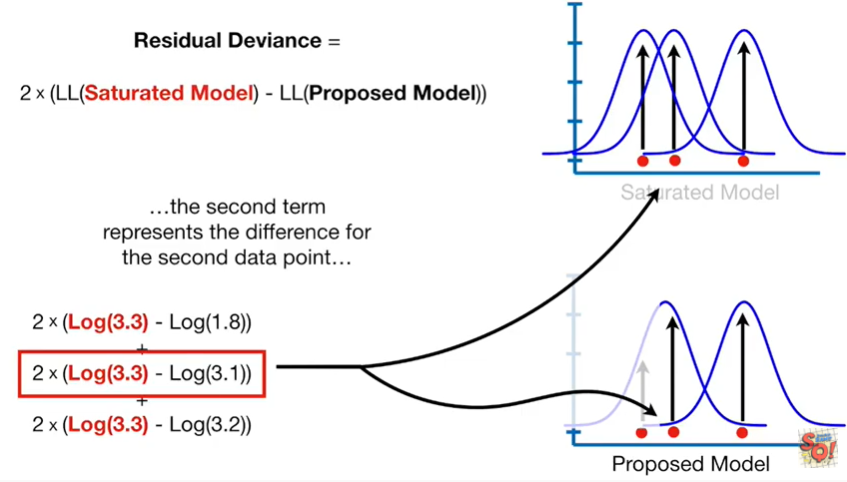
And these terms represent the difference for the third data point.



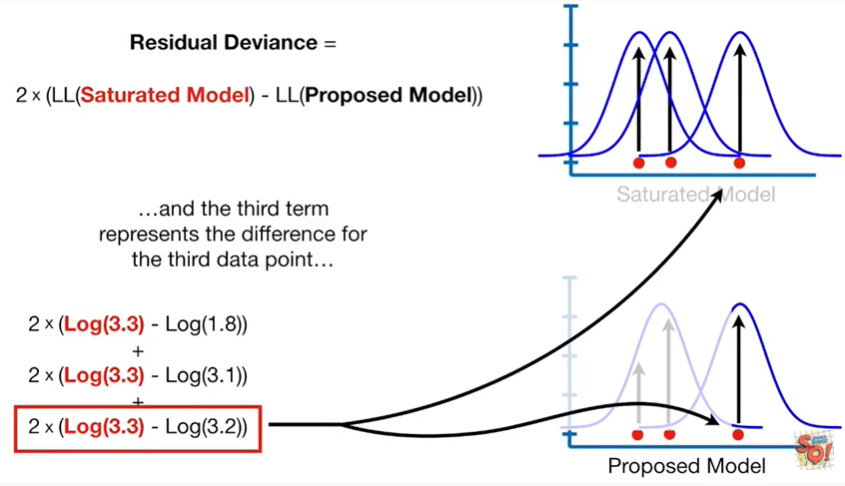
We can rearrange the equation for residual deviance to reflect the difference in likelihoods for the individual data points.



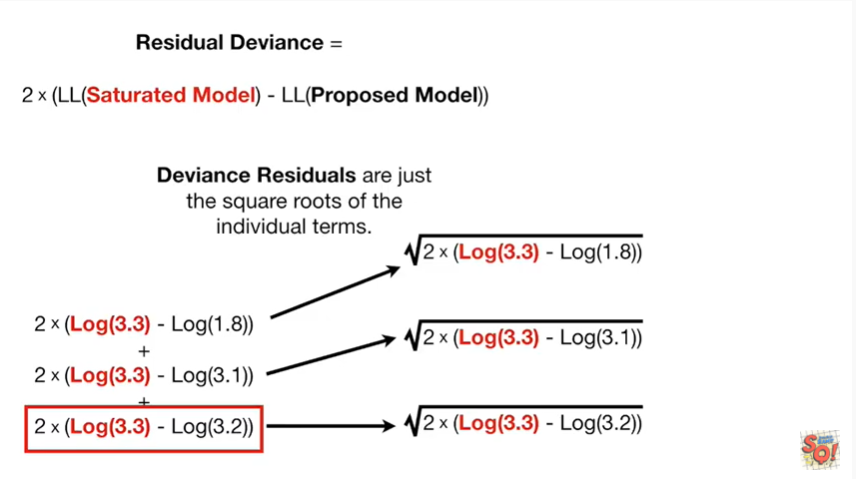
Again the first term represents the difference for the first data point.



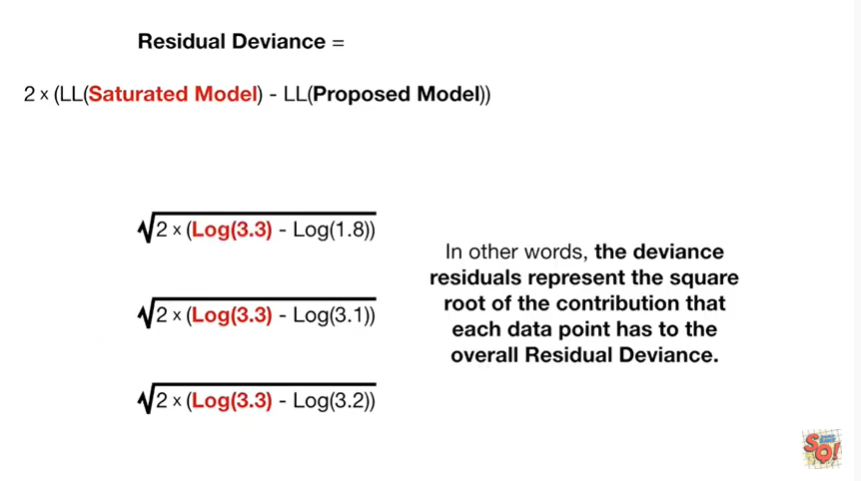
The second term represents the difference for the second data point.



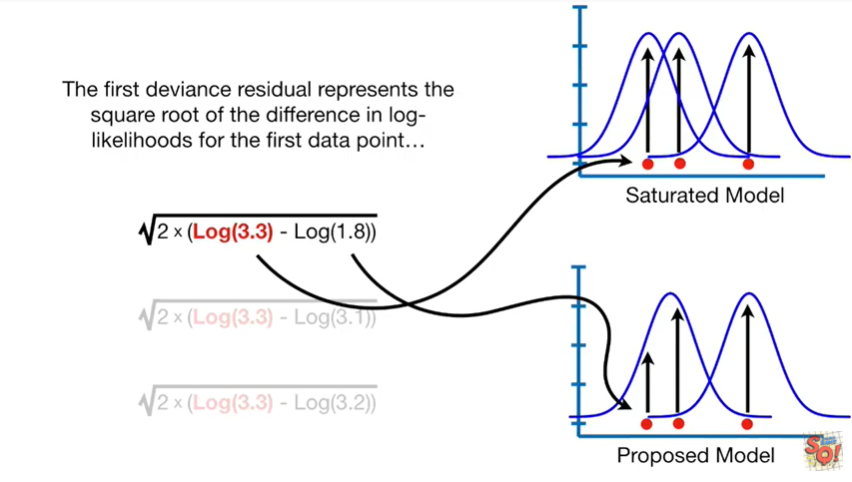
And the third term represents the difference for the third data point.



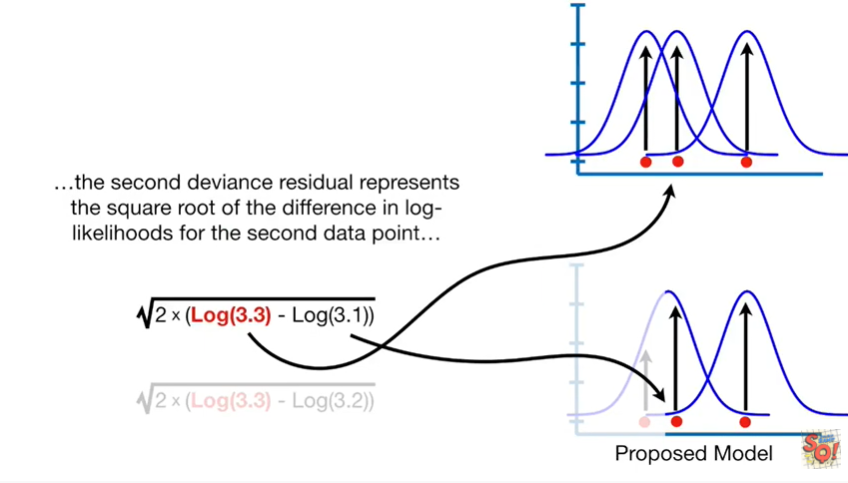
Deviance residuals are just the square roots of the individual terms.



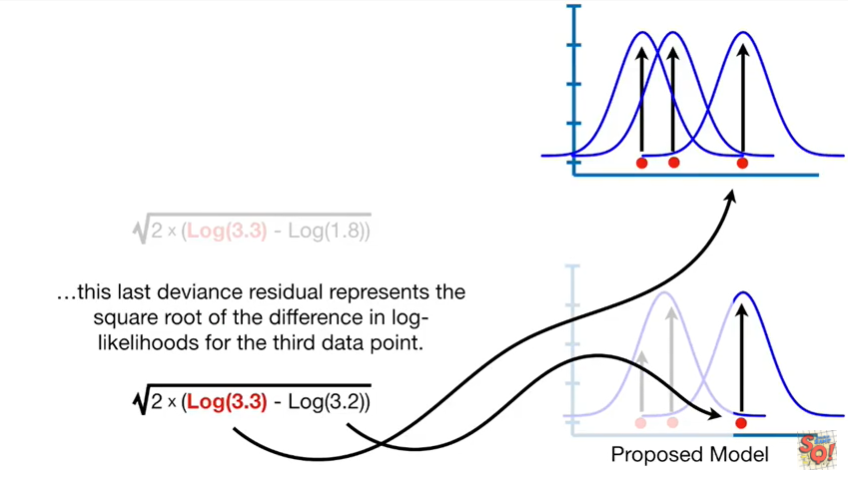
In other words the deviance residuals represent the square root of the contribution that each data point has to the overall residual deviance.



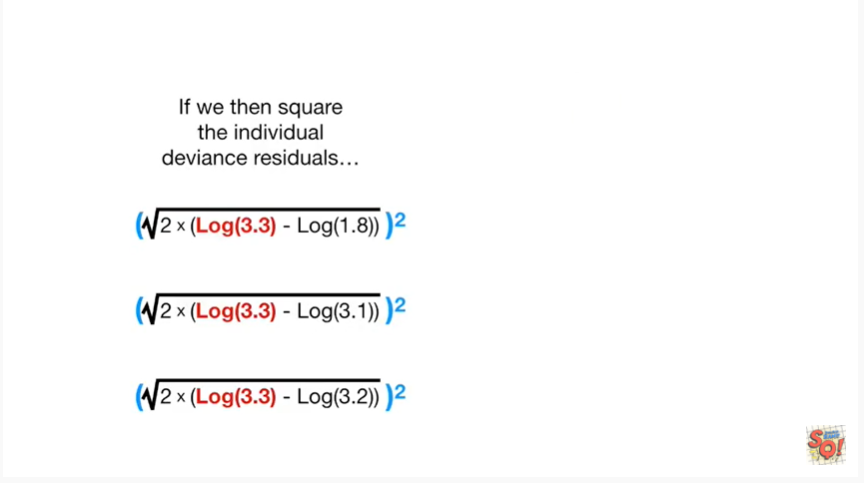
The first deviance residual represents the square root of the difference in log likelihoods for the first data point.



The second deviance residual represents the square root of the difference and log likelihoods for the second data point.

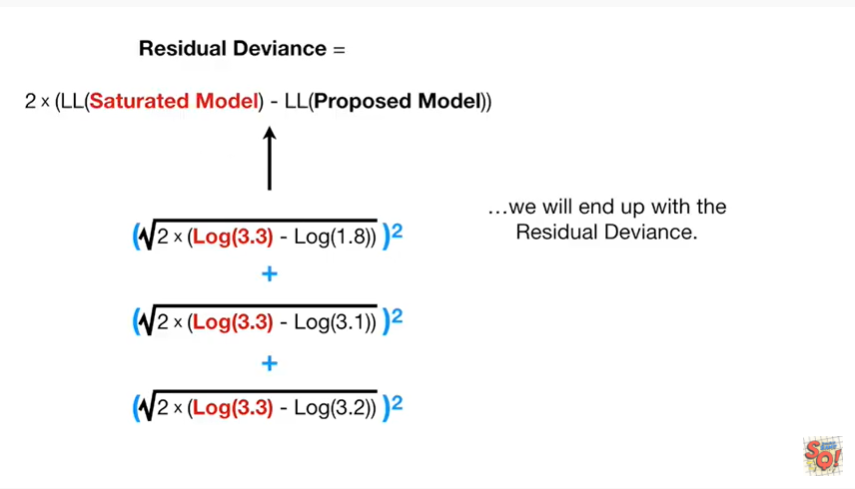


This last deviance residual represents the square root of the difference in log likelihoods for the third data point.

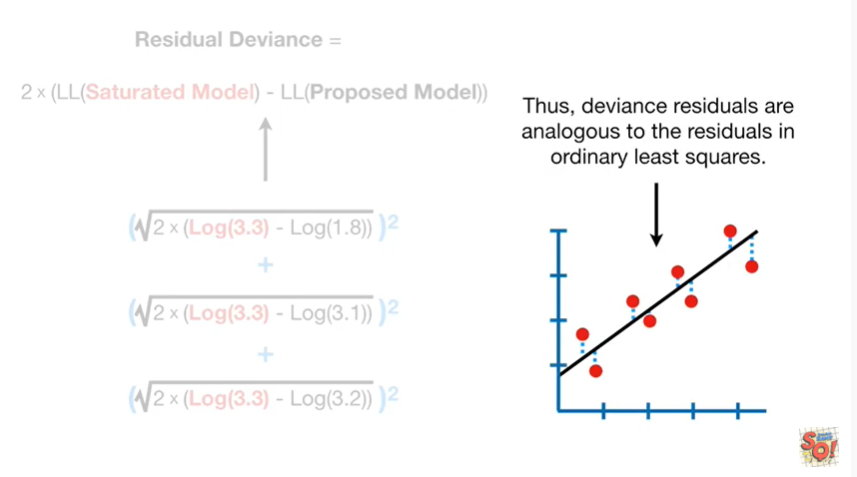


If we then Square the individual deviance residuals

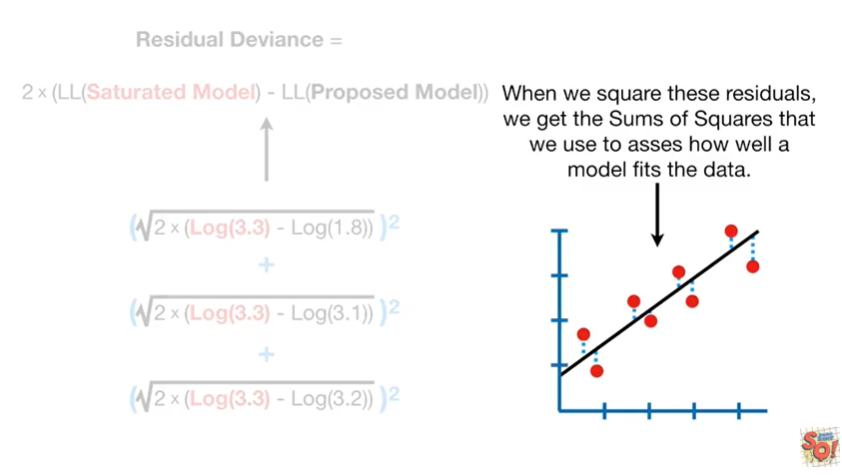
and then add them up



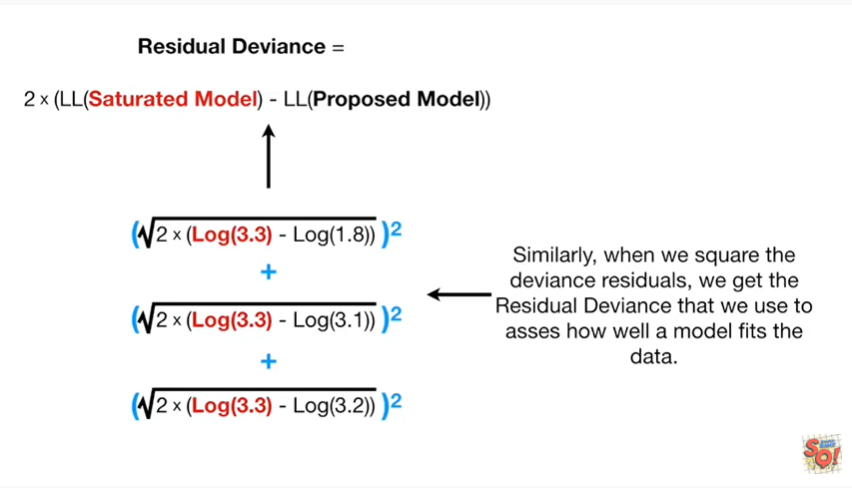
we will end up with the residual deviance.



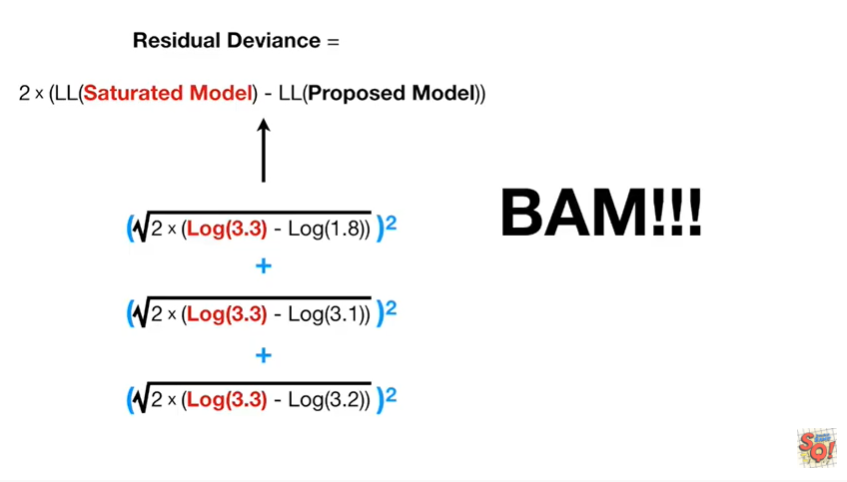
Thus the deviance residuals are analogous to the residuals in ordinary least squares.



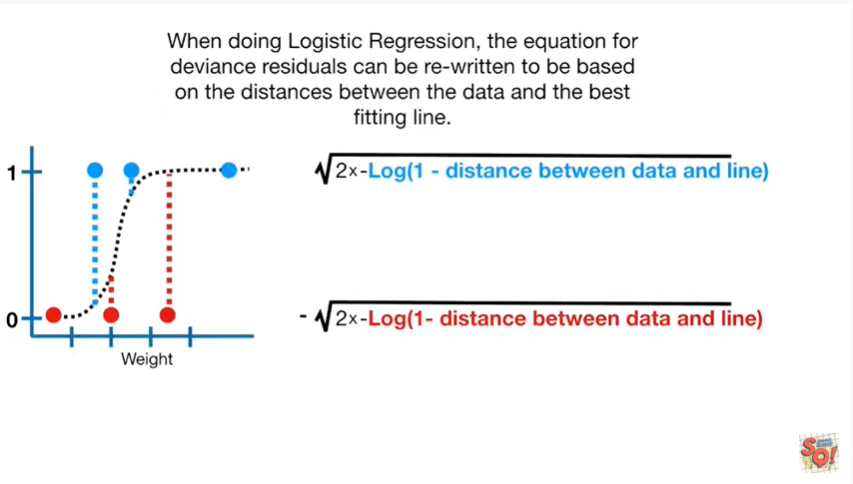
When we square these residuals we get the sums of squares that we use to assess how well the model fits the data.



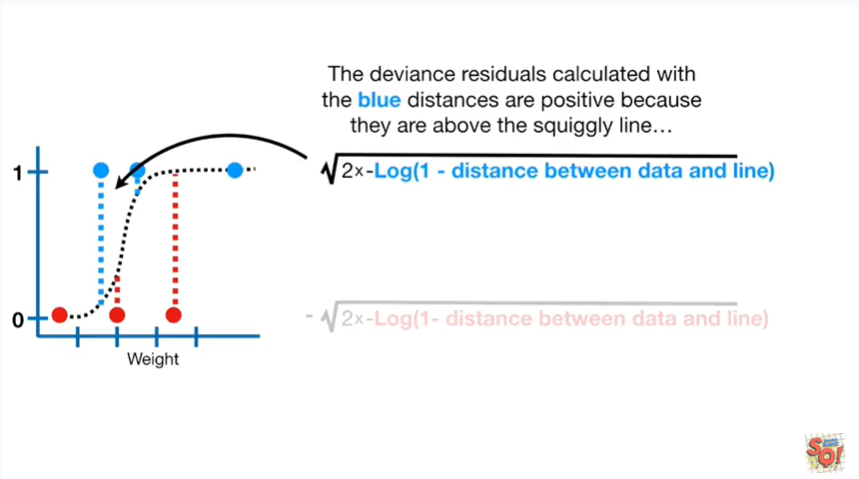
Similarly when we square the deviance residuals we get the residual deviance that we use to assess how well a model fits the data.



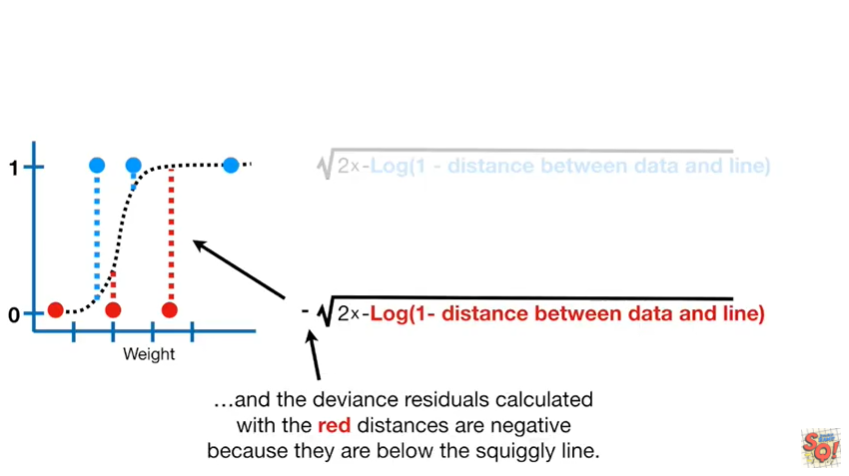
BAM !!!



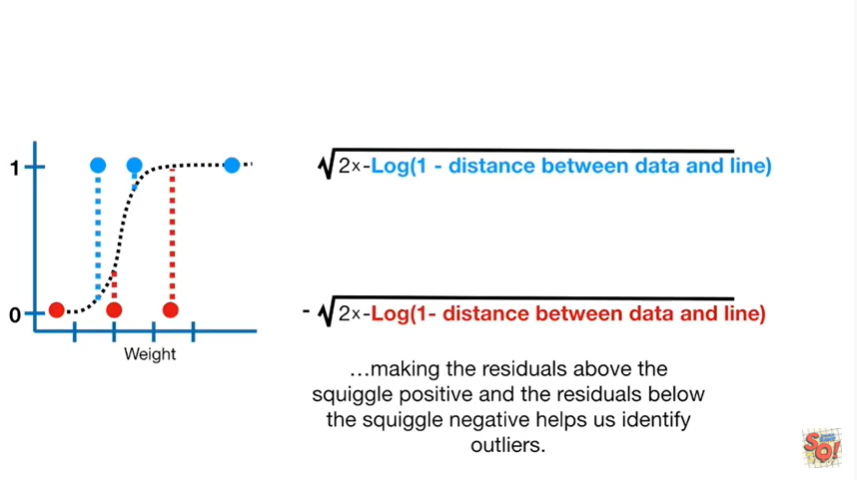
When doing logistic regression the equation for deviance residuals can be rewritten to be based on the distances between the data and the best fitting line.



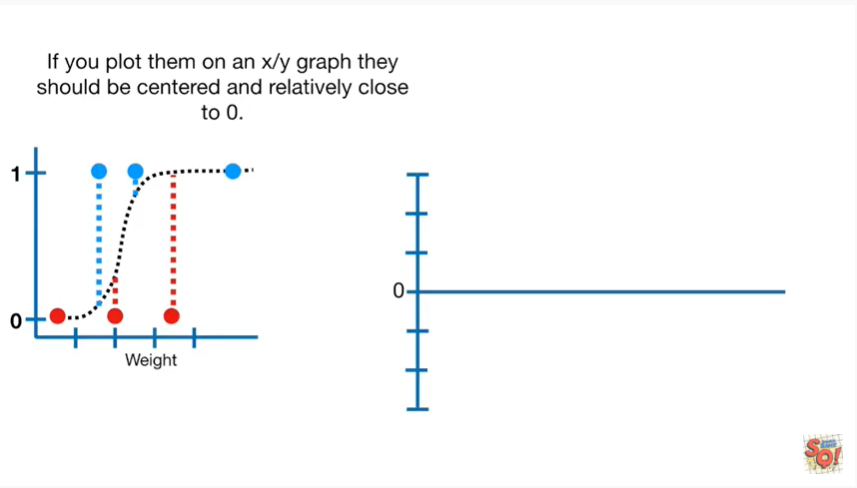
The blue deviance residuals calculated with the blue distances are positive because they are above the squiggly line.



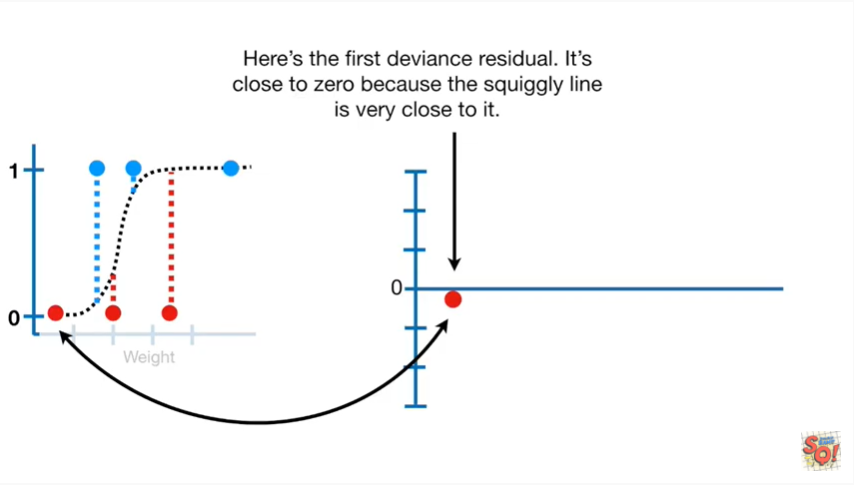
And the deviance residuals calculated with the red distances are negative because they are below the squiggly line.



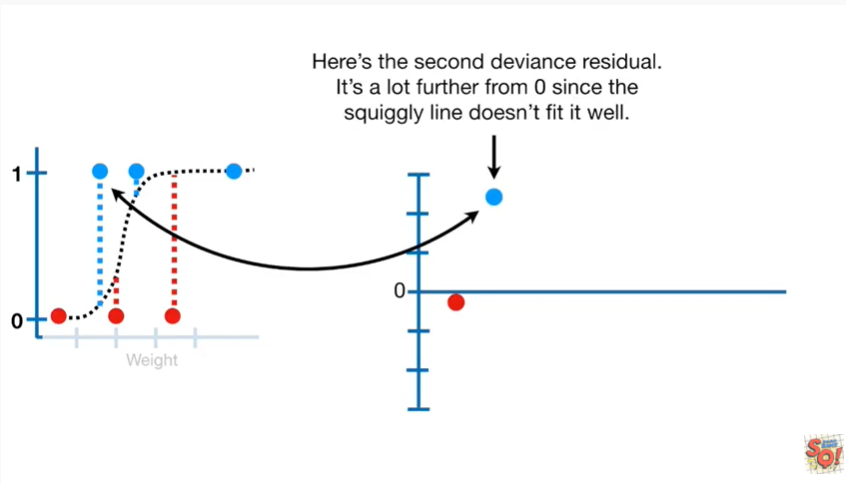
Making the residuals above the squiggle positive and the residuals below the squiggle negative helps us identify outliers.



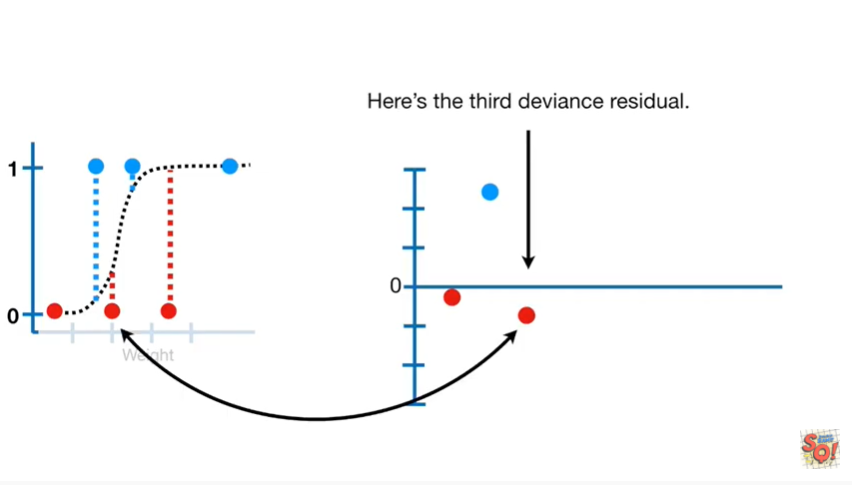
If you plot them on an XY graph they should be centered and relatively close to zero.



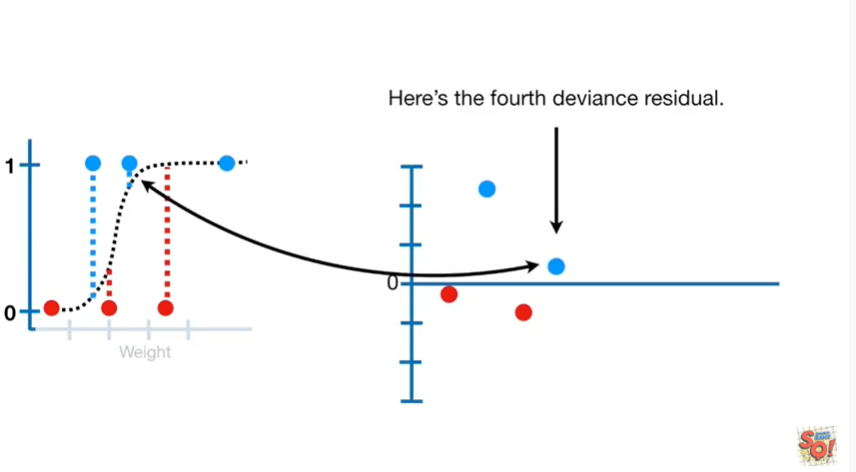
Here's the first deviance residual it's close to zero because the squiggly line is very close to it.



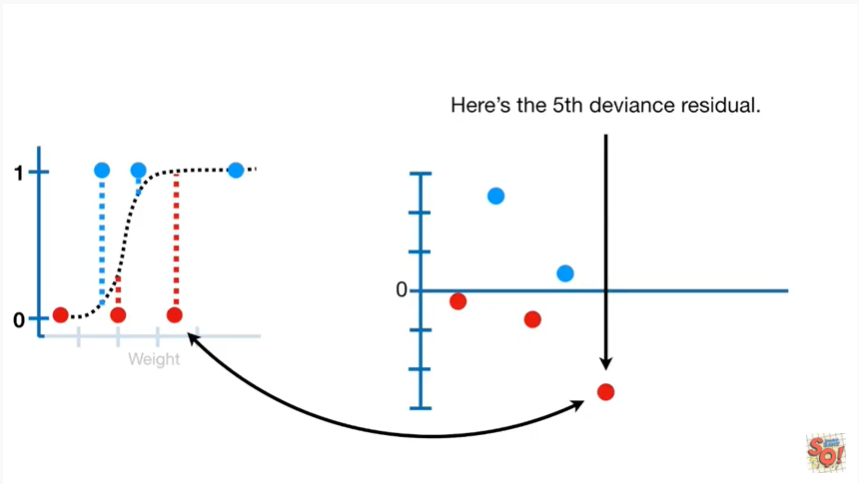
Here's the second deviance residual it's a lot further from zero since the squiggly line doesn't fit it well.



Here's the third deviance residual.



Here's the fourth deviance residual.



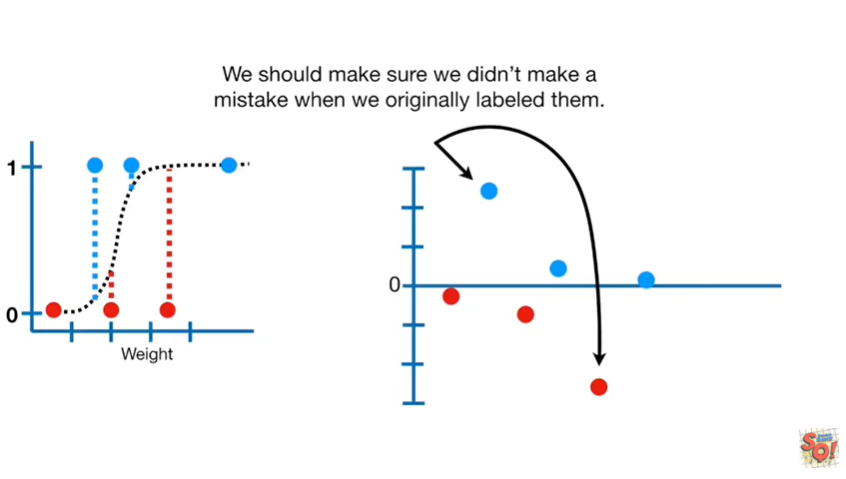
Here's the fifth deviance residual.



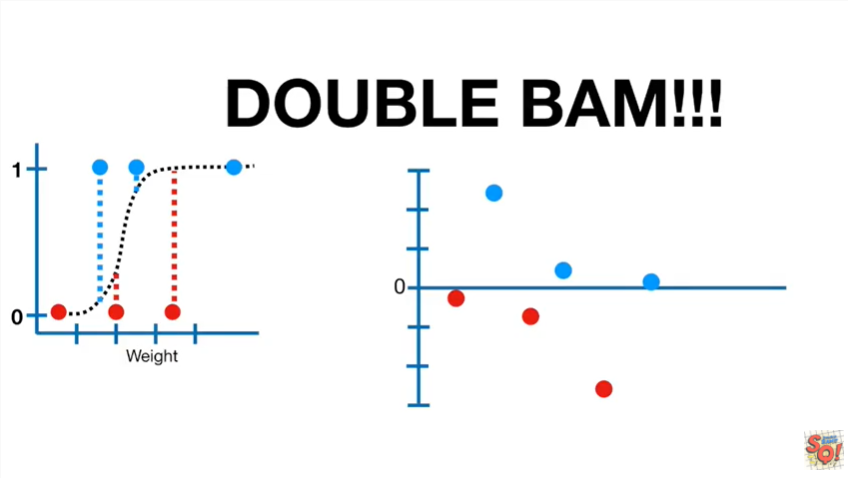
And here's the last deviance residual.



The second and fifth deviance residuals are relatively far from zero and maybe outliers.



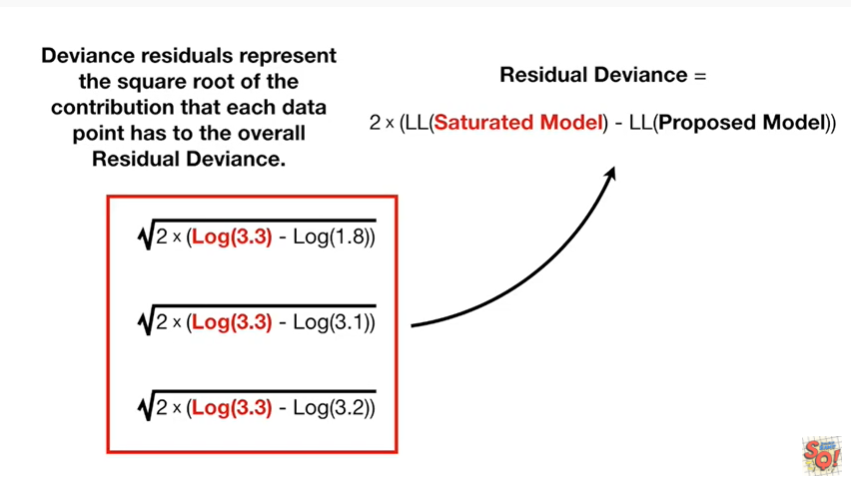
We should make sure we didn't make a mistake when we originally labeled them.



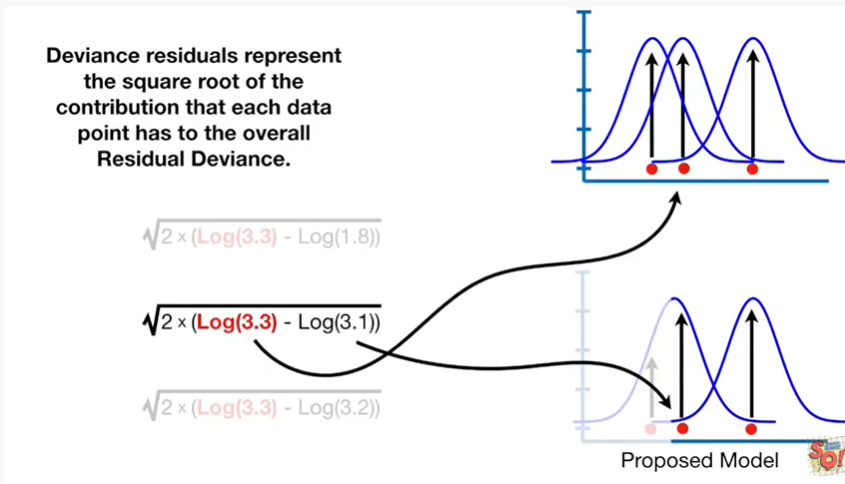
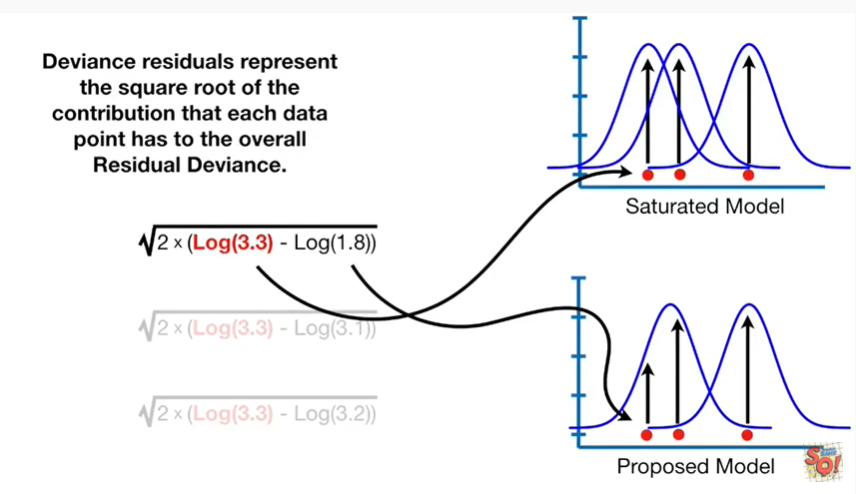
Double BAM.

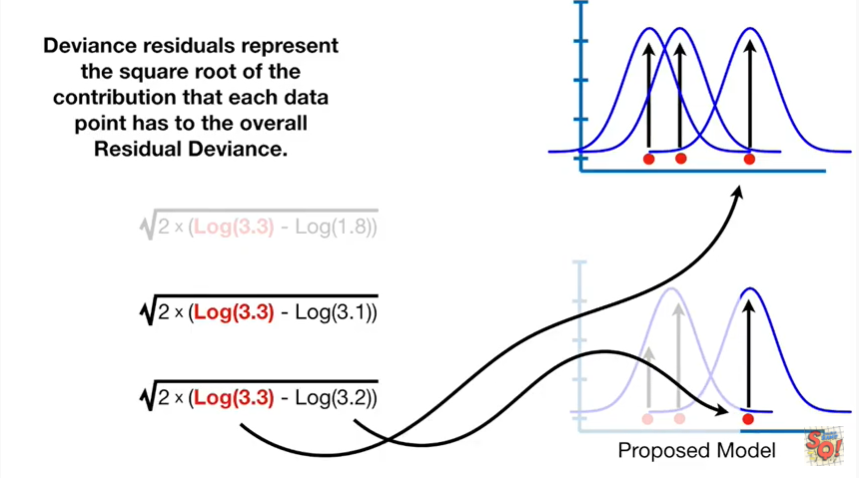


In summary

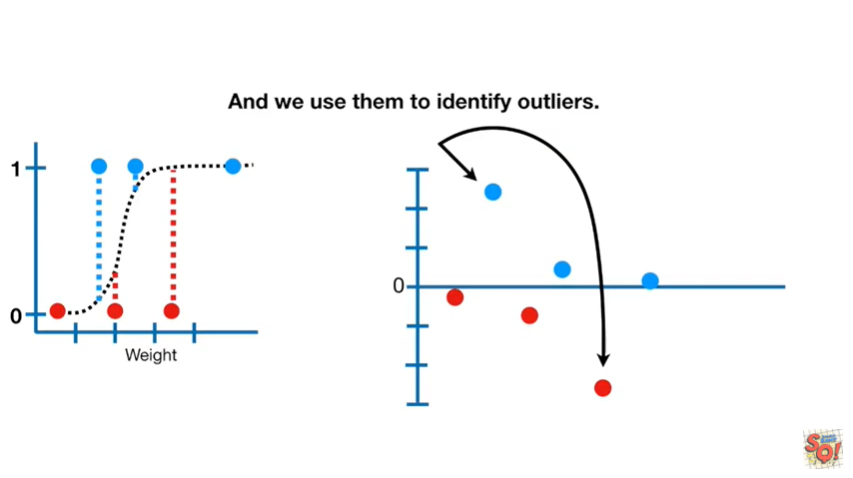


deviance residuals represent the square root of the contribution





that each data point has to the overall residual deviance.



And we use them to identify outliers.