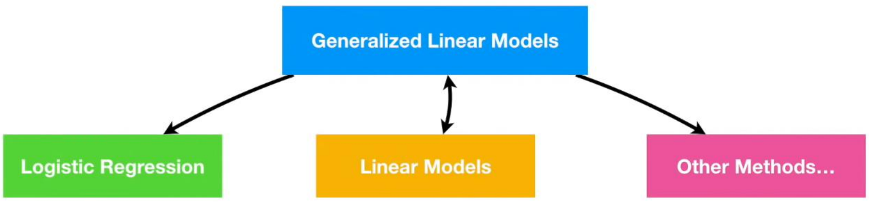
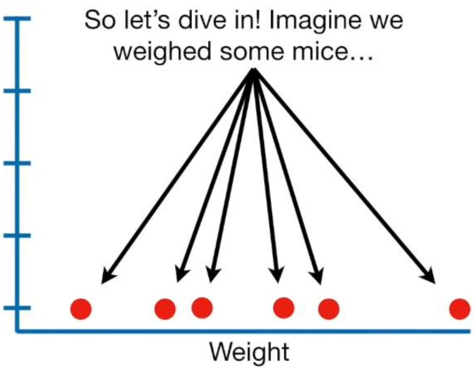
Saturated Models and Deviance Statistics

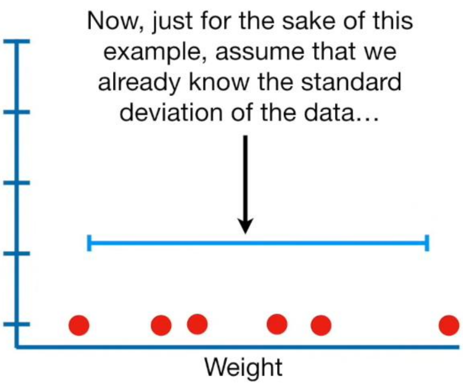
You might have seen it in the context of the log-likelihood based R^2 formula, aka "McFadden's Pseudo R^2"…

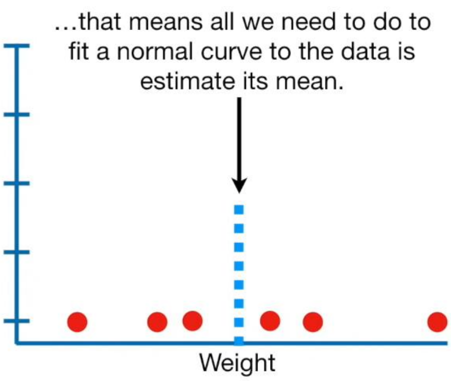
$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(}\textbf{Saturated model}\text{)}}$$

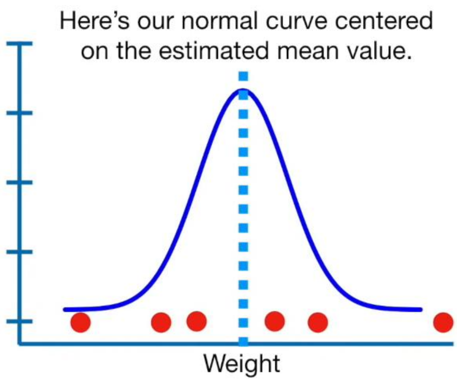
NOTE: The Saturated Model and Deviance Statistics are part of Generalized Linear Models and are intended to be used in a very wide variety of situations.



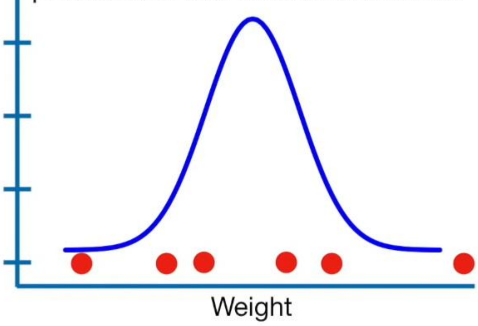




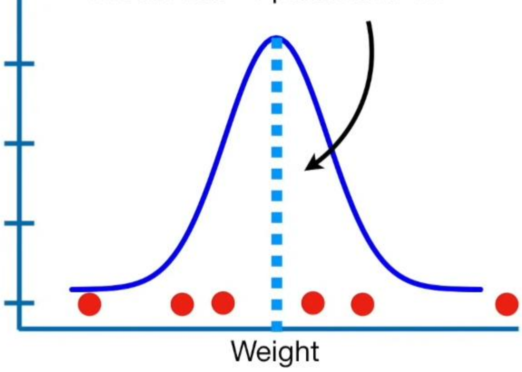




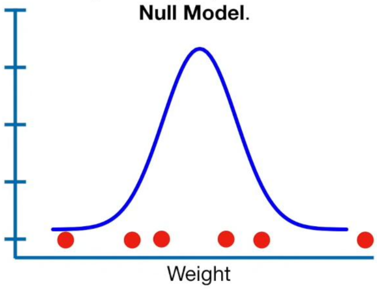
The normal curve is a model of the data. We can use it, instead of the original data, to estimate probabilities and do statistical tests.



Since we only had to estimate the mean for this model, we can say that model has "1 parameter"…



…and since a 1 parameter model is as simple as it gets, we'll call it the **Null Model**.

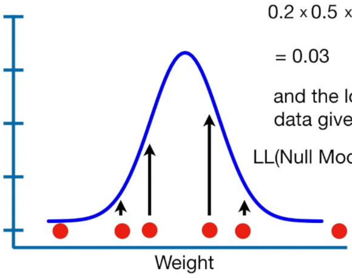


The likelihood of the data, given the Null Model =
$0.2 \times 0.5 \times 1.5 \times 2.5 \times 0.8 \times 0.1$

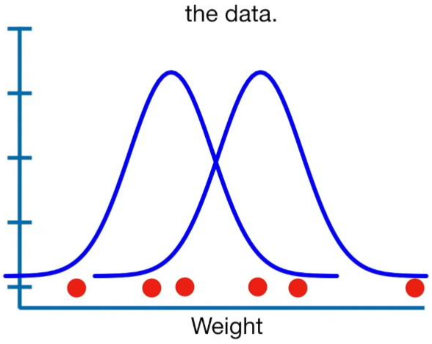
$= 0.03$

and the log-likelihood of the data given the Null Model…

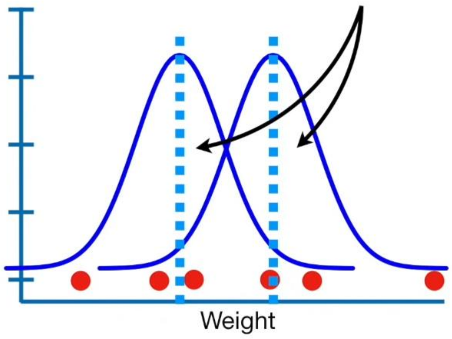
LL(Null Model) = log(0.03) = -3.51



We could also create a slightly fancier model by fitting two normal curves to the data.

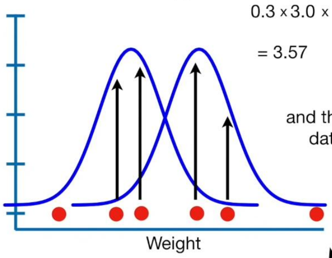


This model would have 2 parameters, one for each mean that we estimated.



The likelihood of the data, given the fancier model =
$0.3 \times 3.0 \times 3.1 \times 3.2 \times 2.0 \times 0.2$

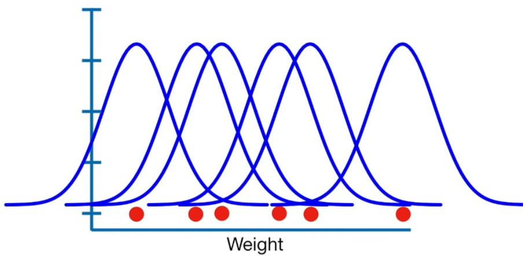
$= 3.57$



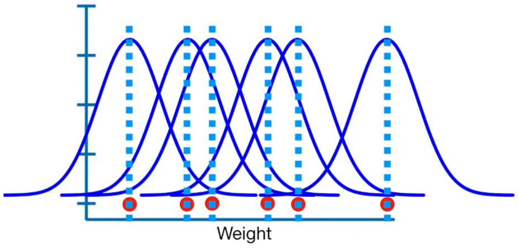
and the log-likelihood of the data given the Fancier Model…

LL(Fancy Model) = log(3.57) = 1.27

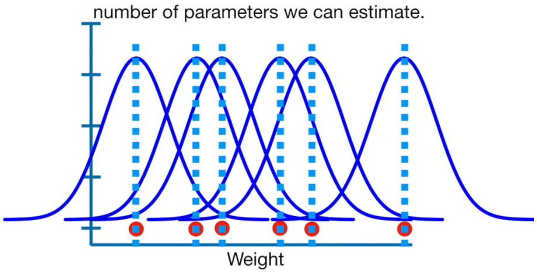
We could also create a super fancy model for our data!!!



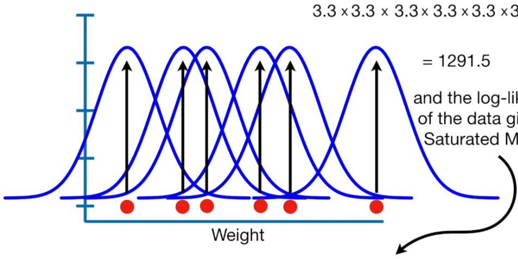
The super fancy model has one parameter per data point.



The super fancy model is called the **Saturated Model** because it maxes out the number of parameters we can estimate.



The likelihood of the data, given the Saturated Model =
3.3 x 3.3 x 3.3 x 3.3 x 3.3 x 3.3

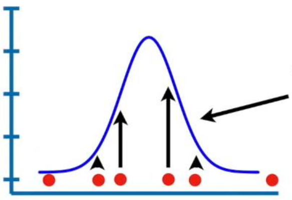


= 1291.5

and the log-likelihood of the data given the Saturated Model…

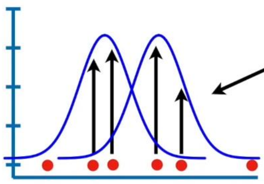
LL(Saturated Model) = log(1291.5) = 7.16

So far we've looked at the likelihood of the data using 3 different models.



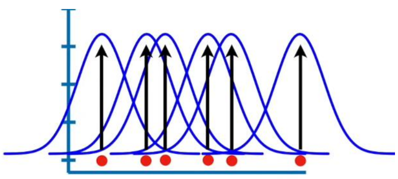
We looked at the likelihood of the data given the **Null Model**, which has only 1 parameter…

Null Model
(only one parameter)



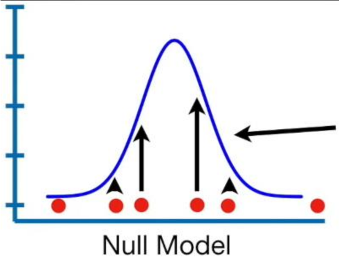
…we looked at the likelihood of the data given a model with 2 parameters.

Let's call this the **Proposed Model**. Imagine it's the one we're really interested in using.



Saturated Model
(one parameter per data point)

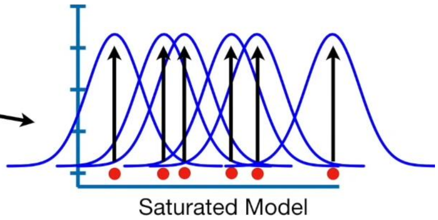
…and we looked at the likelihood of the data given the **Saturated Model**, which has 1 parameter per data point…

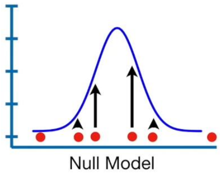


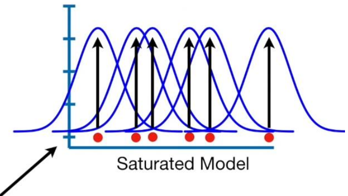
By calculating the likelihood of the **Null Model**, we have a sense of the *worst* case scenario…

…and by calculating the likelihood of the **Saturated Model**, we have a sense of the *best* case scenario.

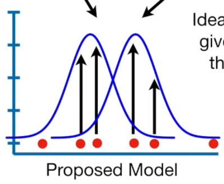
The likelihood of the Saturated Model is a high as it can be.







Ideally, we want the likelihood of the data given our **Proposed Model** to be larger then the **Null Model** and close to the **Saturated Model**…



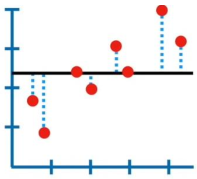
When using likelihoods, we use the **Null and Saturated models** to determine whether the **Proposed Model** does a good job with the data.

In other words, we use the **Null and Saturated Models** to calculate R^2 and its *p*-value for the **Proposed Model**.

Let's start by explaining the role that the Saturated Model plays in the log-likelihood based R^2

$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}}$$

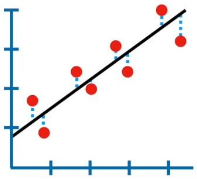
Null Model



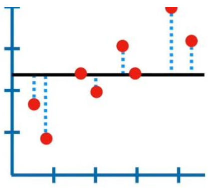
But before we get started, let me review one concept from the "normal" R^2, the one from linear regression calculated with the sums of squares of the residuals.

$$R^2 = \frac{\text{SS(Null Model) - SS(Proposed Model)}}{\text{SS(Null Model)}}$$

Proposed Model



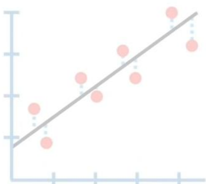
Null Model



With the "normal" R^2, the SS(Null Model) determines the boundary for a bad fit…

$$R^2 = \frac{\text{SS(Null Model) - SS(Proposed Model)}}{\text{SS(Null Model)}}$$

Proposed Model

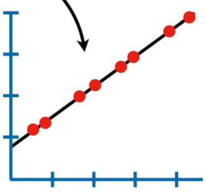


Null Model

…and when the model fits the data perfectly, the residuals are all 0 and SS(Proposed Model) = 0…

$$R^2 = \frac{\text{SS(Null Model) - SS(Proposed Model)}}{\text{SS(Null Model)}}$$

Proposed Model



…in other words, there is a fixed value, 0, that represents a boundary for the best a model can do…

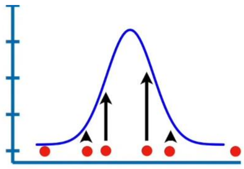
…we get $R^2 = 1$

$$R^2 = \frac{\text{SS(Null Model)}}{\text{SS(Null Model)}} = 1$$

To summarize, the "normal" R^2 has a boundary for how poorly a model can perform, SS(Null Model), and a fixed value, 0, that represents the best possible model.

$$R^2 = \frac{\text{SS(Null Model) - SS(Proposed Model)}}{\text{SS(Null Model)}}$$

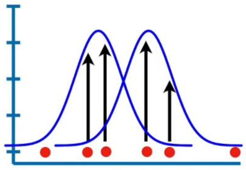
Null Model



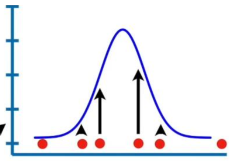
Now let's go back to talking about the log-likelihood based R^2.

$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}}$$

Proposed Model



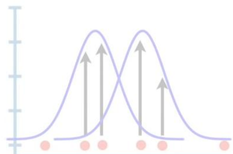
Null Model



Just like before, the Null Model provides a boundary for how poor a model can perform.

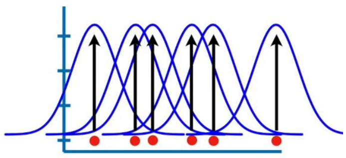
$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}}$$

Proposed Model



However, unlike the traditional R^2, there is no fixed value for an "ideal fit" that works in every situation…

$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}}$$

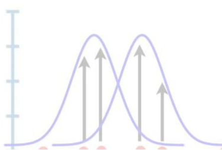


Saturated Model

…and that's where the Saturated Model comes in. The Saturated Model provides an upper bound for what an ideal fit is.

$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}}$$

Proposed Model



Thus, if the Proposed Model is just as good as the Saturated Model, then LL(Proposed Model) = LL(Saturated Model)

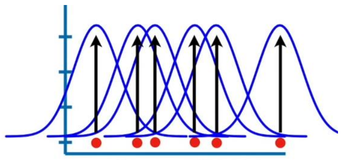
…and that makes the numerator the same as the denominator and R^2 = 1.

$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}} = 1$$

In summary, the log-likelihood of the Saturated Model ensures that the R^2 value ranges between 0 and 1.

Now let's talk about **Deviance**, a concept related to the Saturated Model that will, ultimately, lead us to a *p*-value for the R^2.

We're going to talk about two types of Deviance **Residual Deviance** and **Null Deviance**.

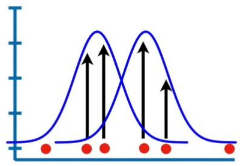


Saturated Model

Residual Deviance is defined as:

2 x (LL(**Saturated Model**) - LL(**Proposed Model**))

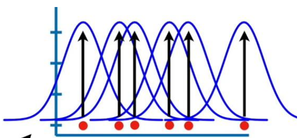
The "2 times" part makes the difference in these log-likelihoods have a Chi-squared distribution with degrees freedom equal to the difference in the number of parameters.



Proposed Model

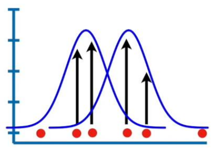
The **Residual Deviance** is defined as:

2 x (LL(**Saturated Model**) - LL(**Proposed Model**))



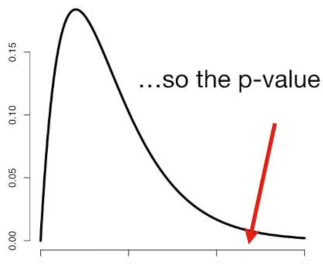
Saturated Model

In this case, the degrees of freedom is 4, the number of parameters in the **Saturated Model**, 6, minus the number of parameters in the **Proposed Model**, 2.

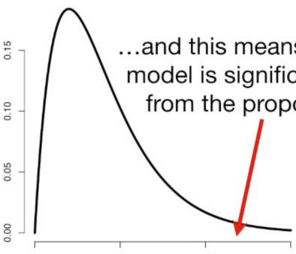


Proposed Model

2 x (7.16 - 1.27) = 11.78



…so the p-value = 0.02…



…and this means the saturated model is significantly different from the proposed model.

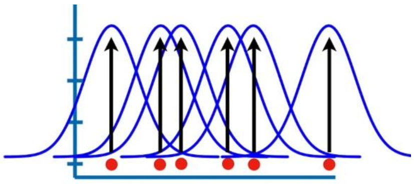
NOTE: For this Chi-square test to work correctly, the **Proposed Model** and **Saturated Model** have to be "nested".

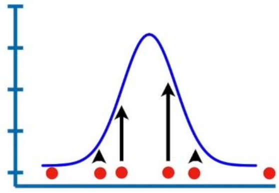
In other words, the Proposed Model has to be a simpler version of the Saturated Model (and not just any old model).

The **Null Deviance** is defined as:

2 × (LL(**Saturated Model**) - LL(**Null Model**))

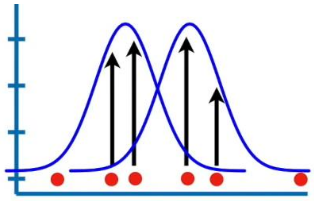
Again, the "2 times" part makes the difference in these log-likelihoods have a Chi-squared distribution with degrees freedom equal to the difference in the number of parameters.





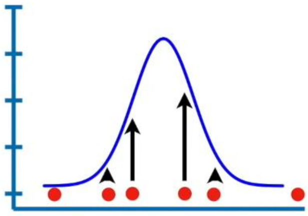
In this case, the degrees of freedom is 5, the number of parameters in the **Saturated Model**, 6, minus the number of parameters in the **Null Model**, 1.

…you can calculate the p-value for the log-likelihood R^2 from the difference in the deviances…



Null Deviance - **Residual Deviance**

…this difference is a Chi-square value with degrees of freedom equal to the difference in the number of parameters for the Proposed Model and the Null Model.



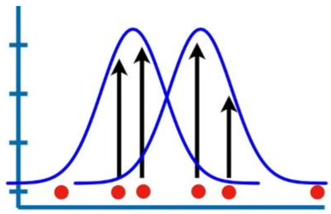
$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - LL(Saturated Model)}}$$

$R^2 = 0.45$

Thus, the p-value for the log-likelihood R^2 value we calculated before is 0.002.

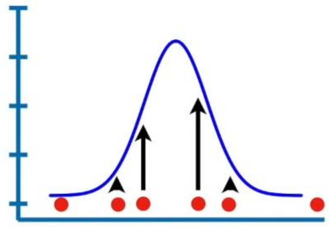
This means that not only does our model have a reasonable effect size (the R^2 value), we know that that value is not by chance.

Alternatively, we can calculate the significance of the R^2 directly by comparing the Proposed Model to the Null Model, since the Null Model is nested within the Proposed Model.



2 x (LL(**Proposed Model**) - LL(**Null Model**))

Again, the "2 times" part makes the difference in these log-likelihoods have a Chi-squared distribution with degrees freedom equal to the difference in the number of parameters.

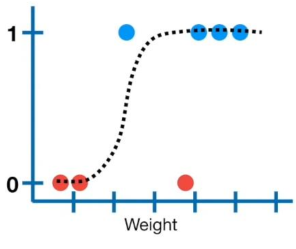


However, I mentioned the method that uses Deviances because the Null and Residual Deviances are very frequently reported. In contrast, the LL(Proposed Model) and the LL(Null Model) are not always reported.

OK, we've talked about the Saturated Model, Residual Deviance and Null Deviance.

Now let's talk about why you can ignore the Saturated Model entirely when you do Logistic Regression.

When you do logistic regression, you're fitting a squiggle to your data and calculating the log-likelihood of the data given that squiggle.

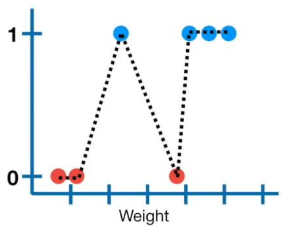


log(likelihood of data given the squiggle) =

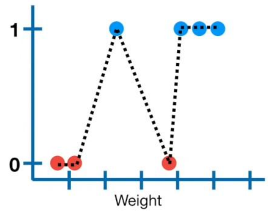
log(0.49) + log(0.91) + log(0.91) + log(0.92) + log(1 - 0.9) + log(1 - 0.01) + log(1 - 0.01)

= -3.3

However, with the saturated model, the squiggle fits the data perfectly…



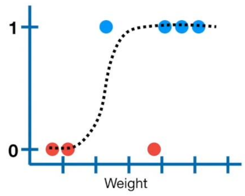
…and the log-likelihood of the data is 0, because log(1) = 0, so the log-likelihood is just a big sum of zeros.



log(likelihood of data given the saturated squiggle) =

log(**1**) + log(**1**) + log(**1**) + log(**1**) +
log(**1 - 0**) + log(**1 - 0**) + log(**1 - 0**)

This means that the equation for the likelihood based R^2 is this…



$$R^2 = \frac{\text{LL(Null Model) - LL(Proposed Model)}}{\text{LL(Null Model) - 0}}$$