Maximum likelihood: a bit of context

Outline

Step back – the bigger picture

From data to maximum likelihood estimation
(and back again):

- Population and sample
- Model choice
- Parameter estimation
- Uncertainty
- Interpretation

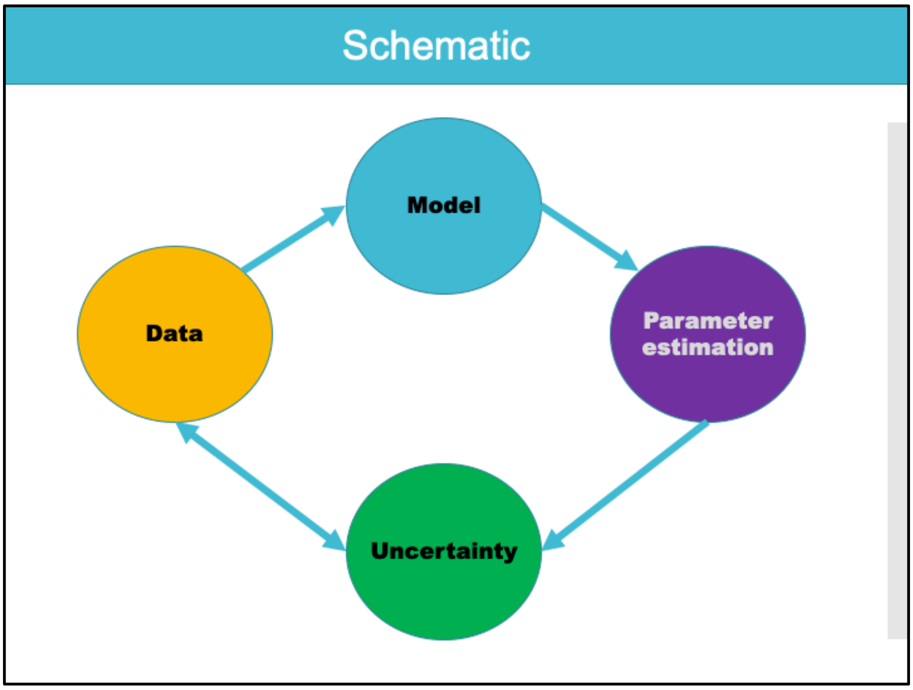
Steps of modelling (week 2 recap)

1. Choose a model for your data
2. Get estimates of the parameters
3. Quantify uncertainty in the estimates
4. Interpret the results

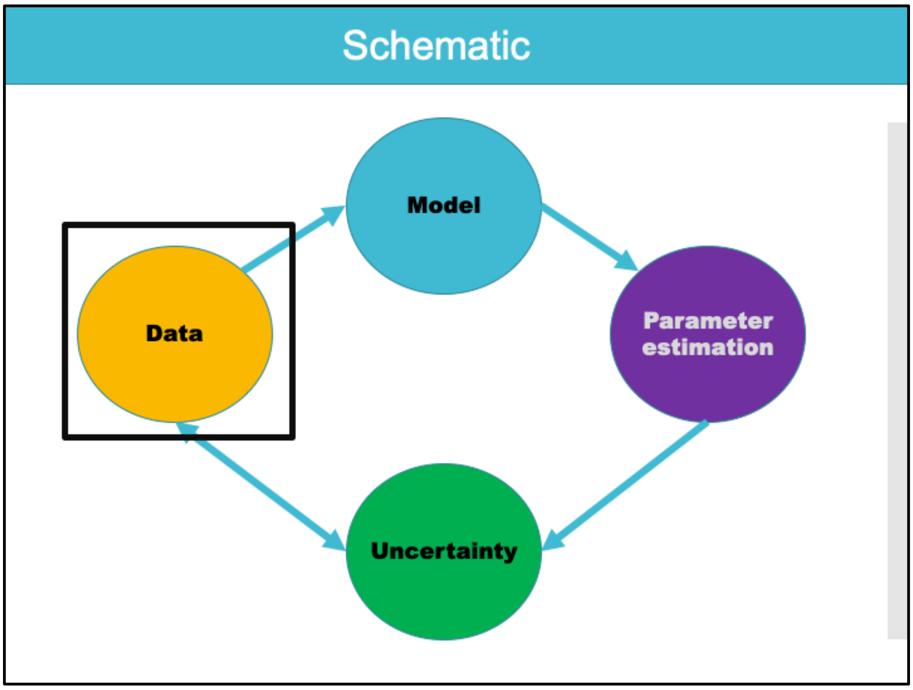
Steps of modelling (week 2 recap)

1. Choose a model for your data
2. Get estimates of the parameters
3. Quantify uncertainty in the estimates
4. Interpret the results

Maximum likelihood



Can think of all of these processes as linking together. Will cover each of these bits in more detail today and explain where maximum likelihood estimation fits in

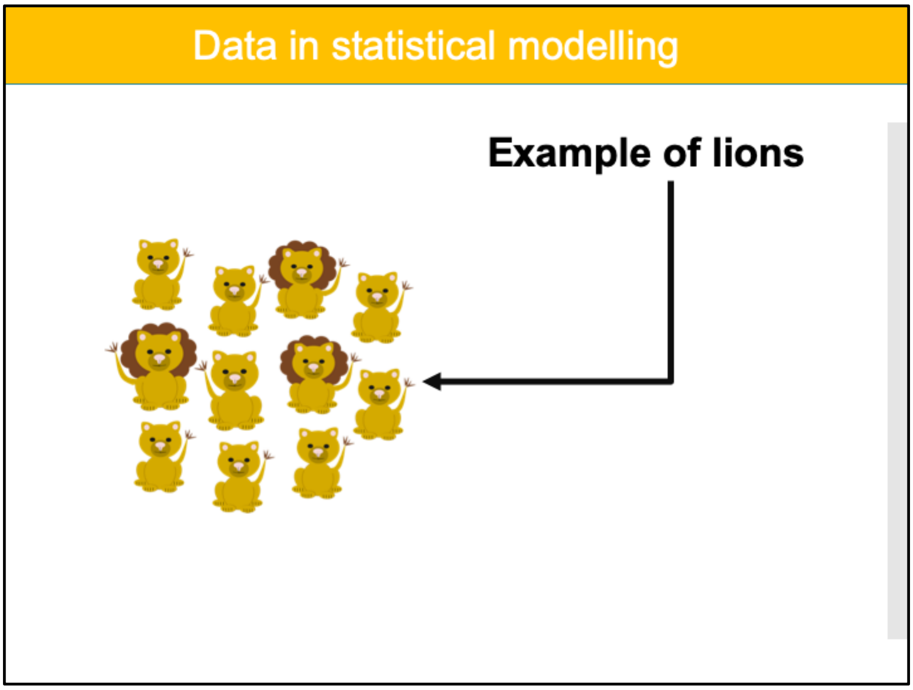


Begin with data

Data in statistical modelling

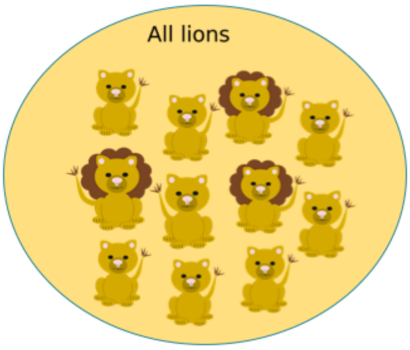
Example of lions

Imagine a group of lions

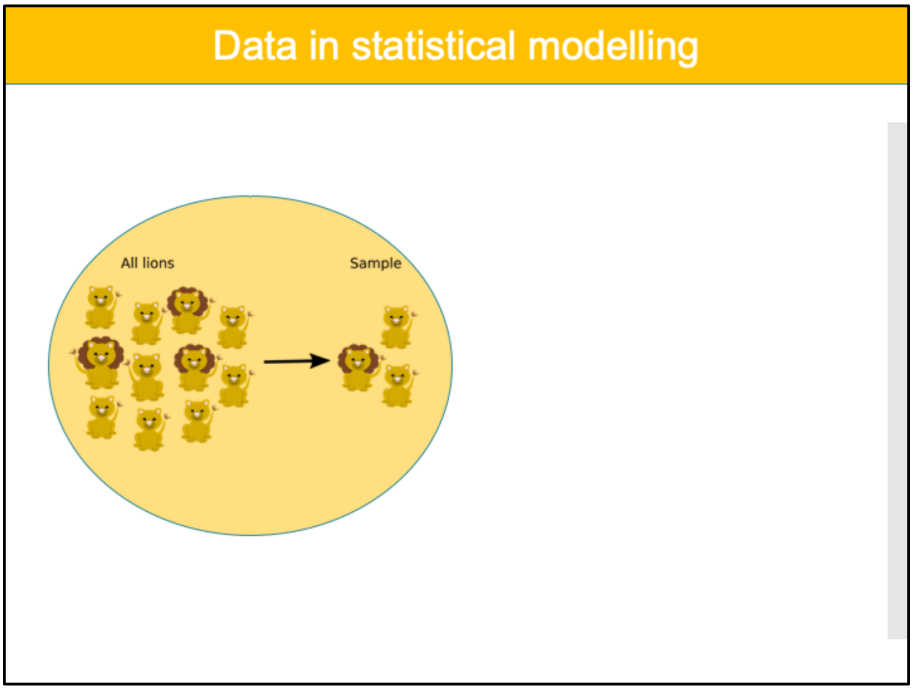


Imagine a group of lions

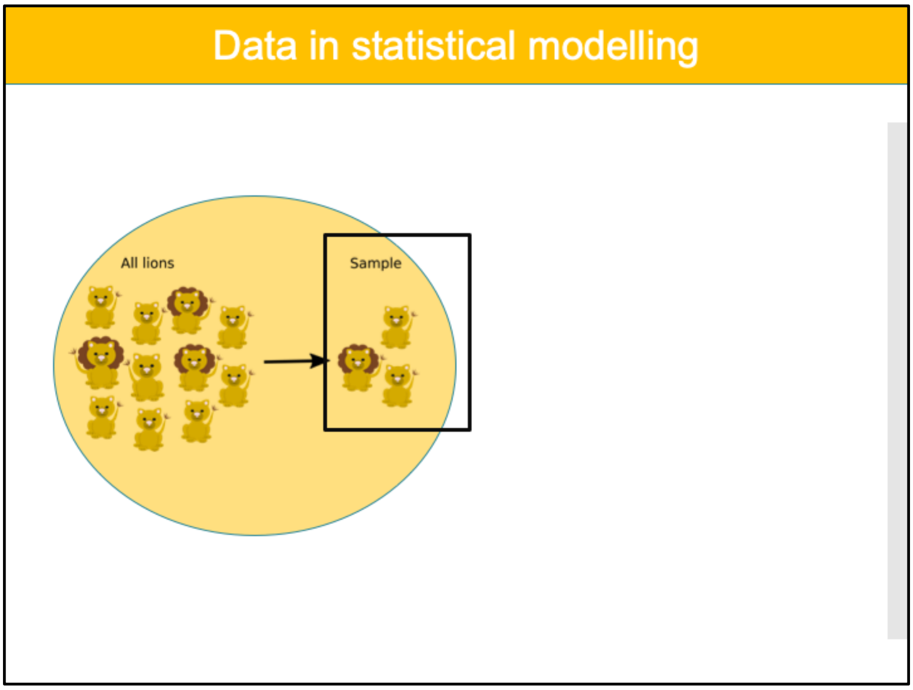
Data in statistical modelling



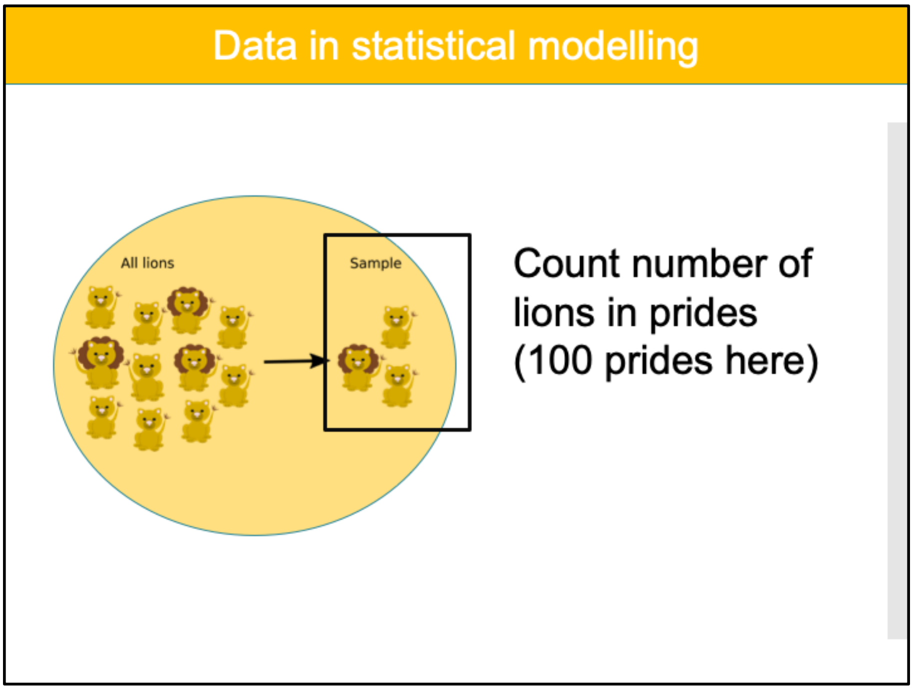
We want to collect data on the lions. We cannot catch them all.



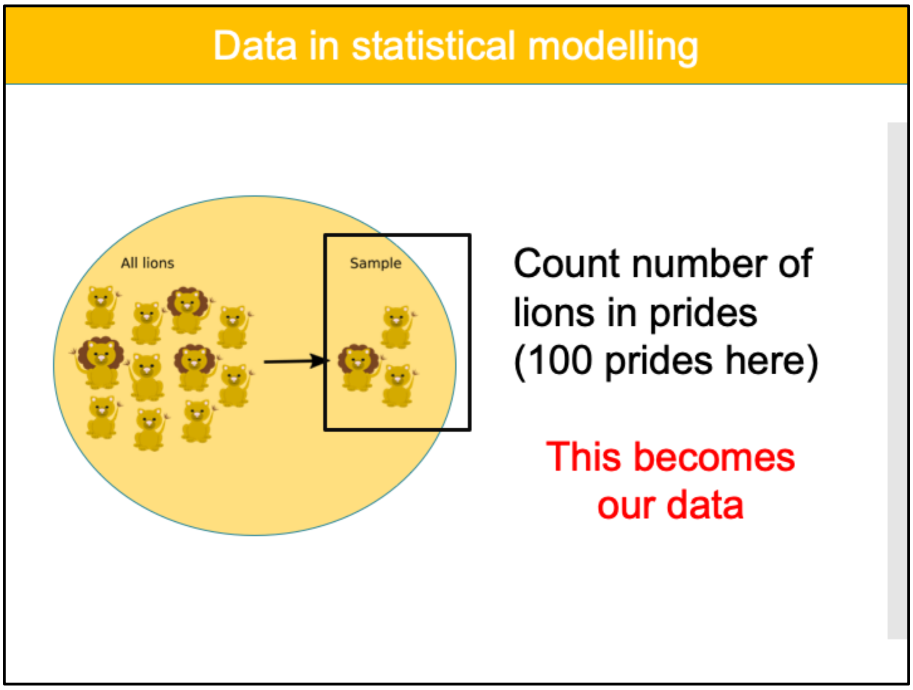
In our sample, count number of lions in the pride



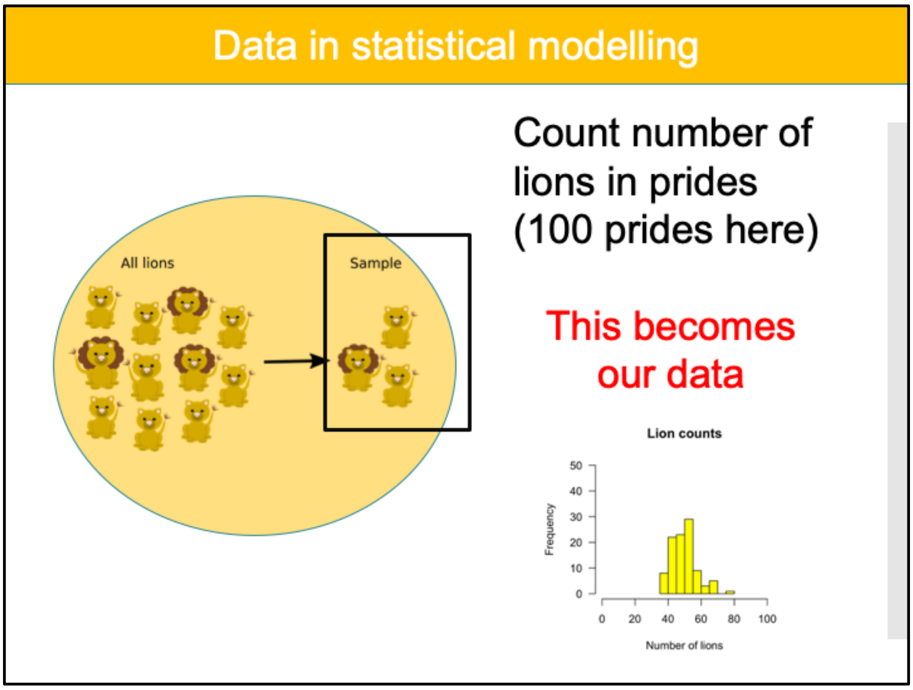
In our sample, count number of lions in the pride



In our sample, count number of lions in the pride

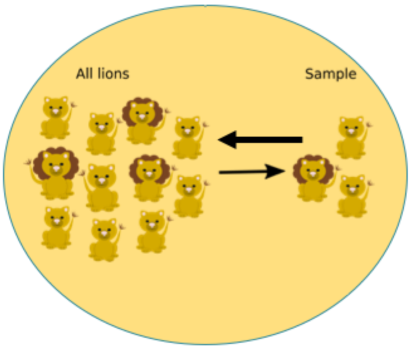


In our sample, count number of lions in the pride



In our sample, count number of lions in the pride

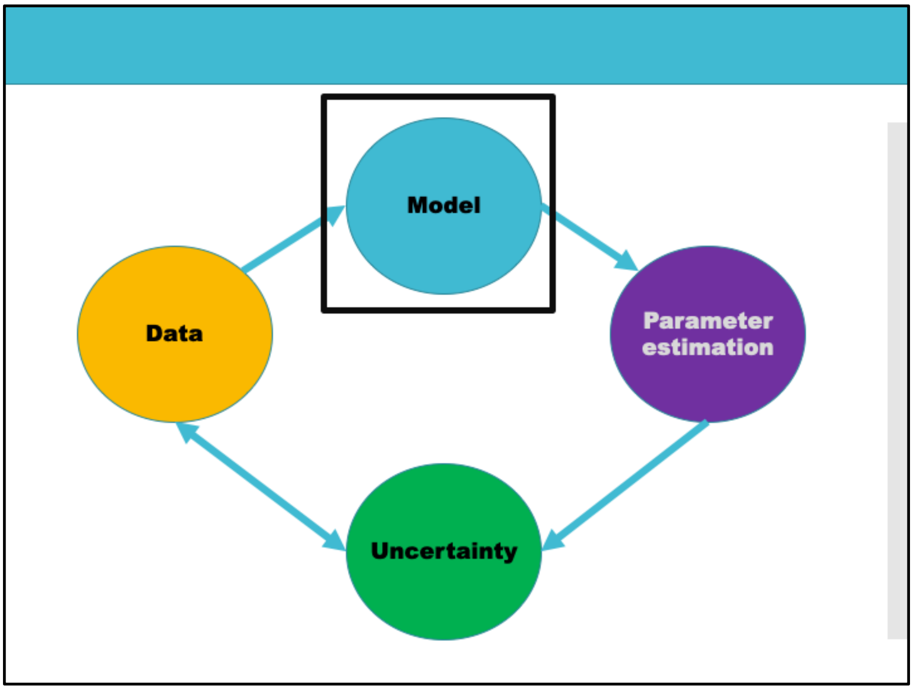
Data in statistical modelling



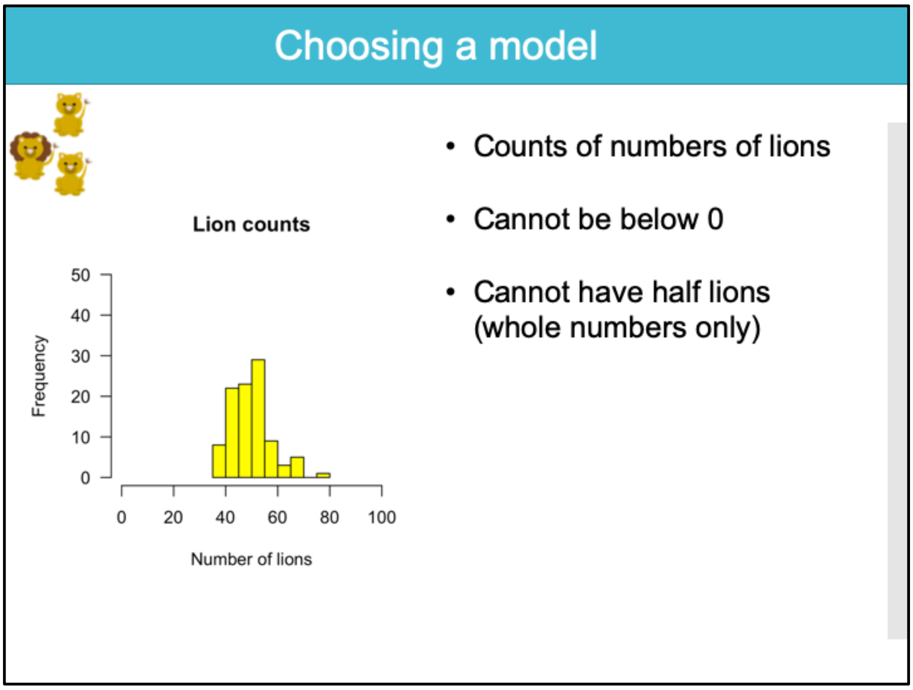
Want to say something about total population of lions

Need a model!

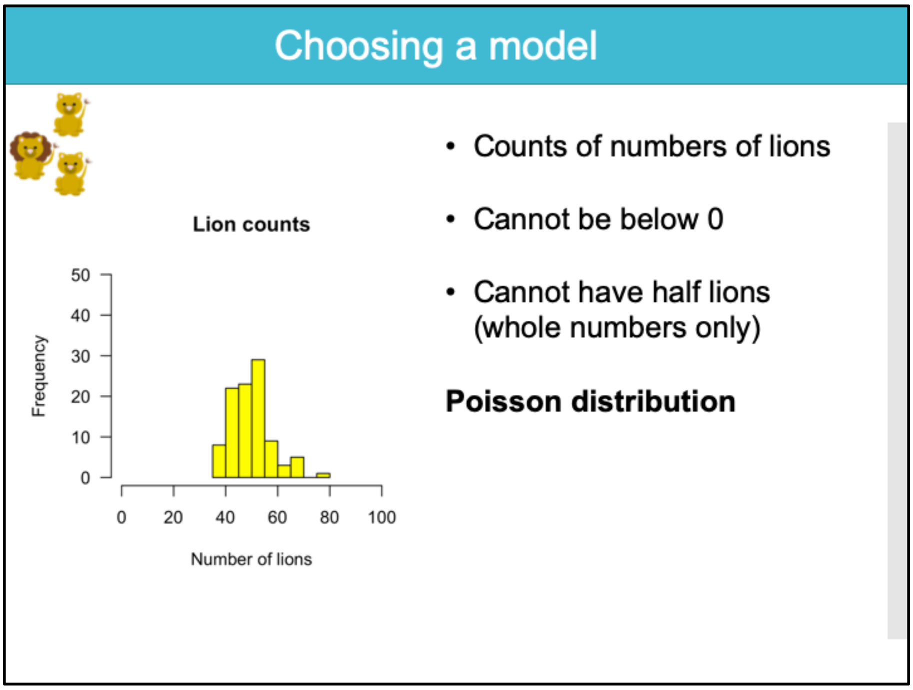
We want to say something about the population because this is more interesting. We don't want to only describe our sample (sometimes you do, but not often), we will have missed some lions in each pride and some whole prides. But we want to give an idea of the number of lions in each pride for all lions. Other examples = say something about all birds in a wood from sampling 100, say something about all trees in a woodland from sampling 50 etc.



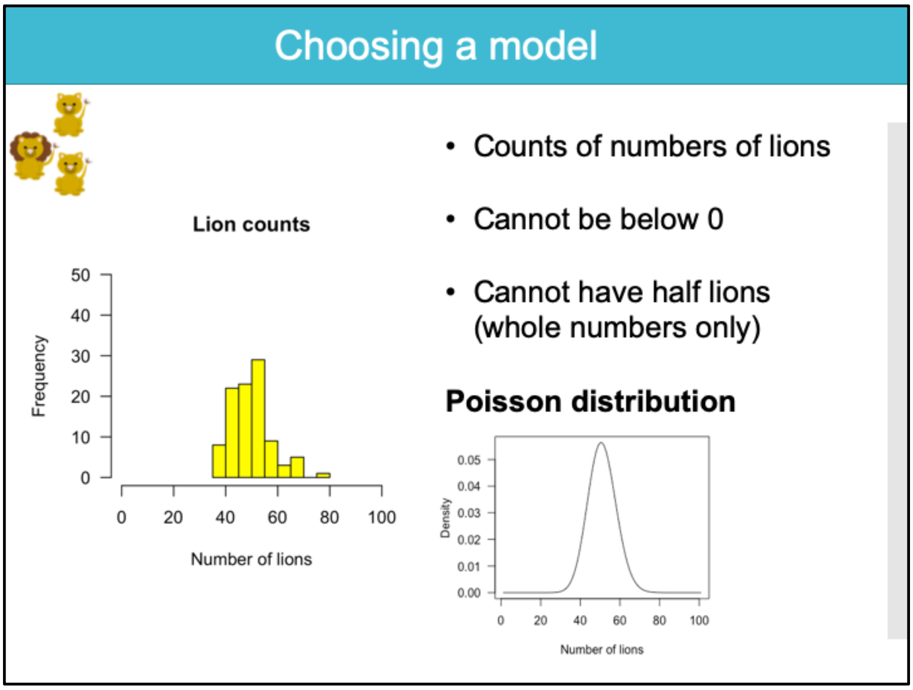
Once we have data, want to choose a model



We have our data and have plotted it for 100 prides. We need to find a model that can represent how these data were generated.



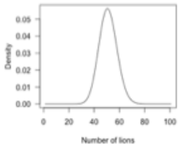
A Poisson distribution fits this characteristics.



You can see it follows a similar shape to our data too.

Any model

Our example = **Poisson distribution**

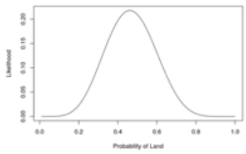


Any model

Our example = Poisson distribution

But works for any model

E.g. Binomial distribution (land and sea)



The same process is true for other datasets and models

Any model

Our example = Poisson distribution

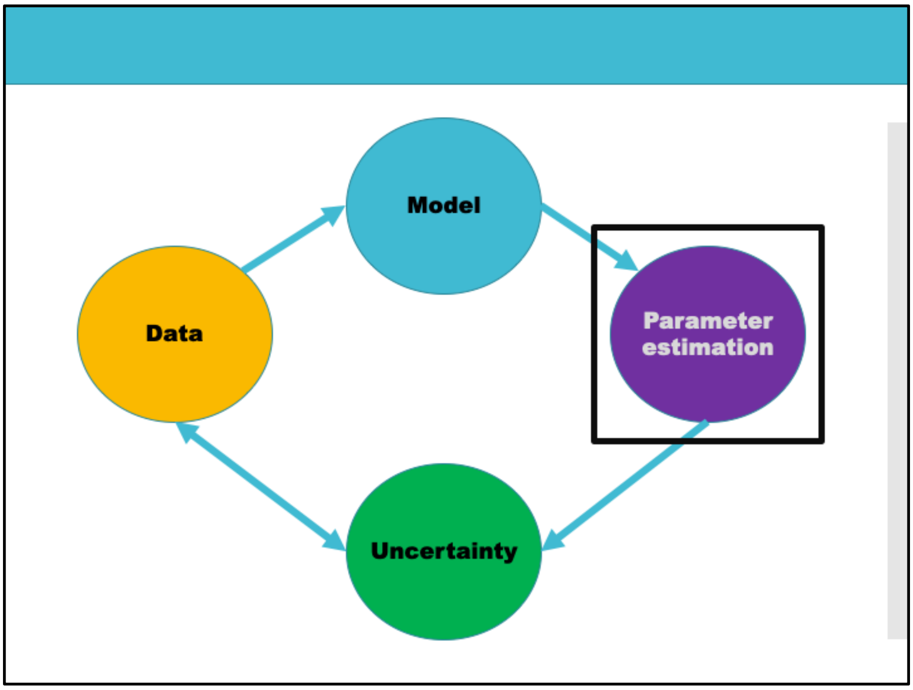
But works for any model

E.g. Binomial distribution (land and sea)

Linear equation (regression – coming soon)

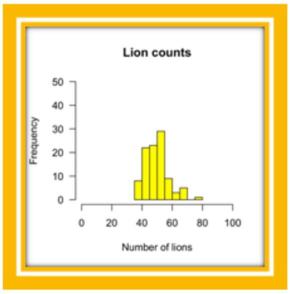
Almost anything

This is GENERAL idea

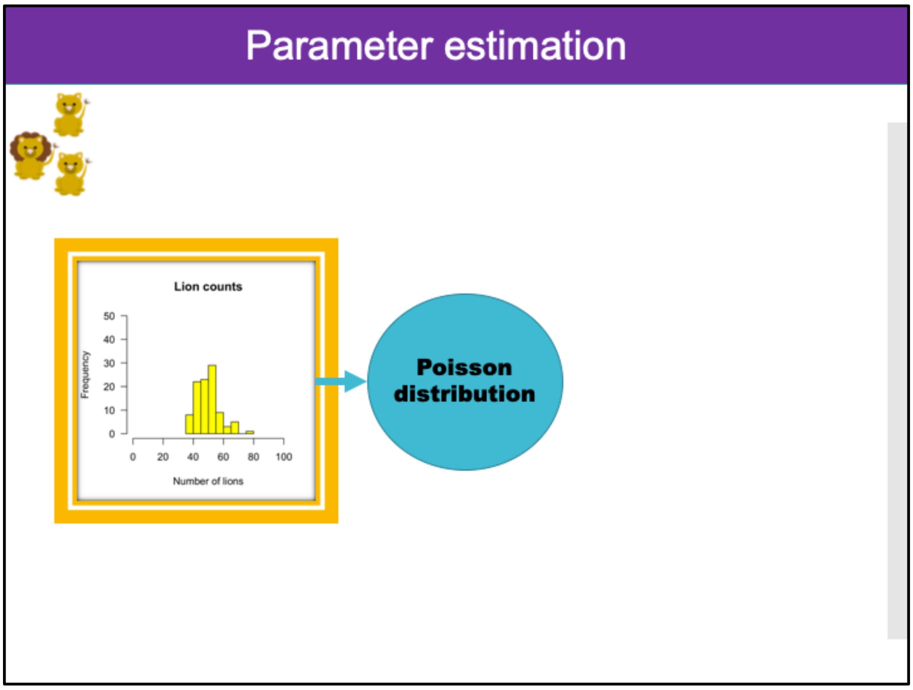


Now we have a model, need to estimate the parameters of the model

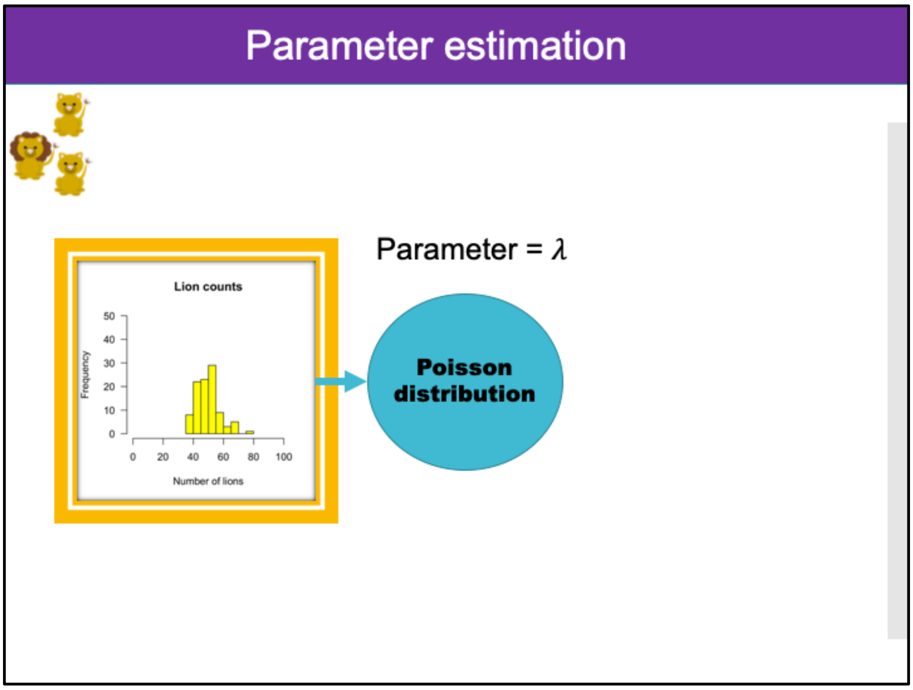
Parameter estimation



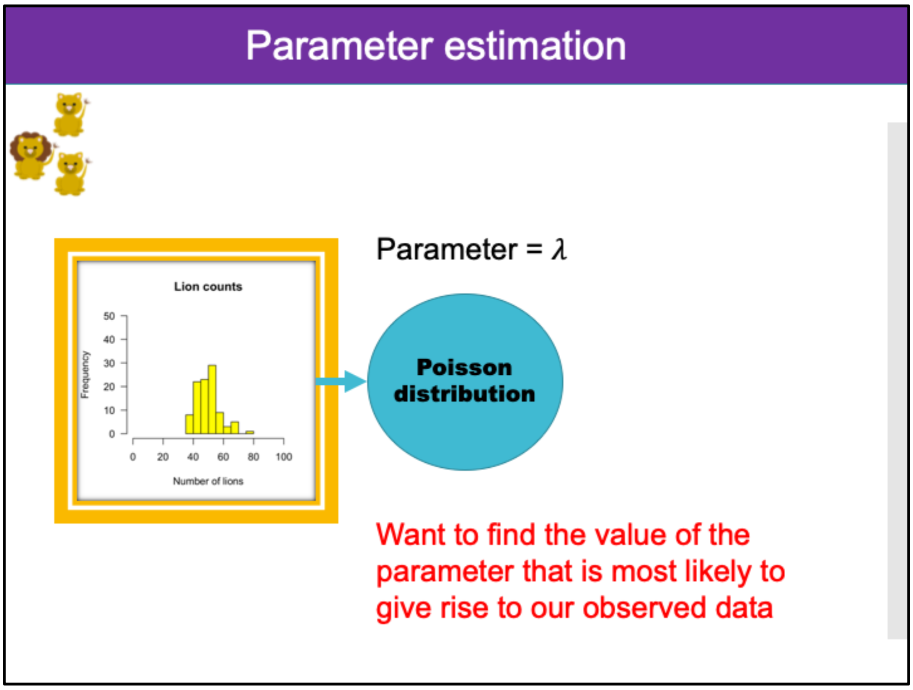
Back to our data, we have these counts of lions and we have plotted them in a histogram



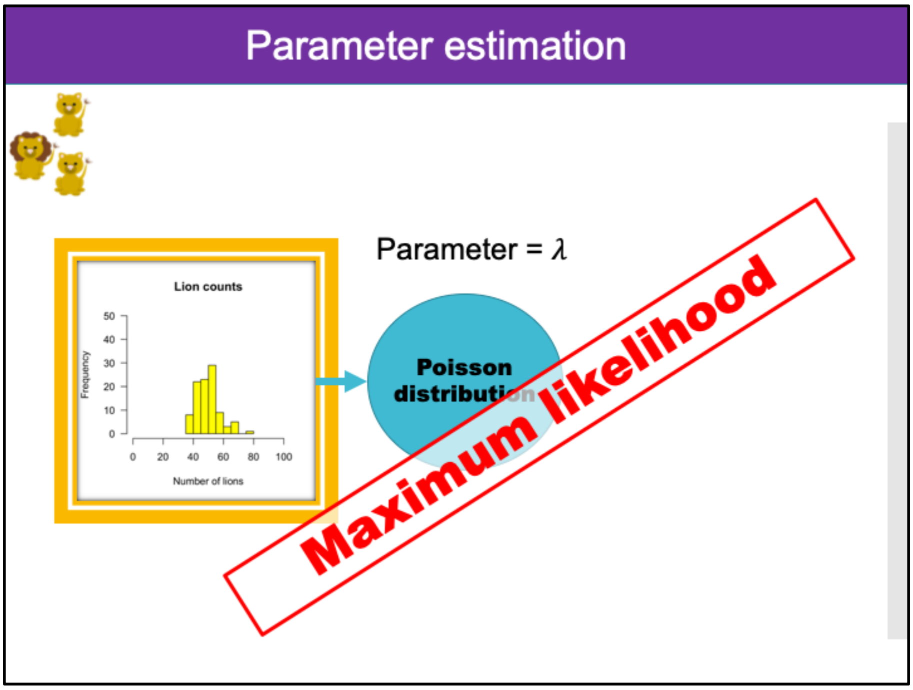
We have also chosen our model



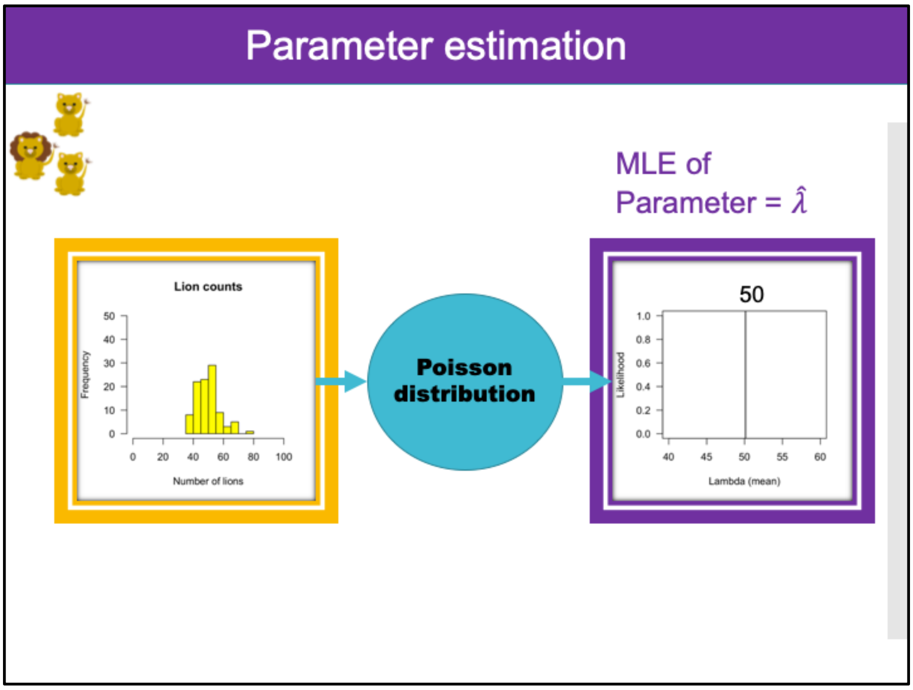
We know that this is characterised by a single parameter lambda (mean and the variance)



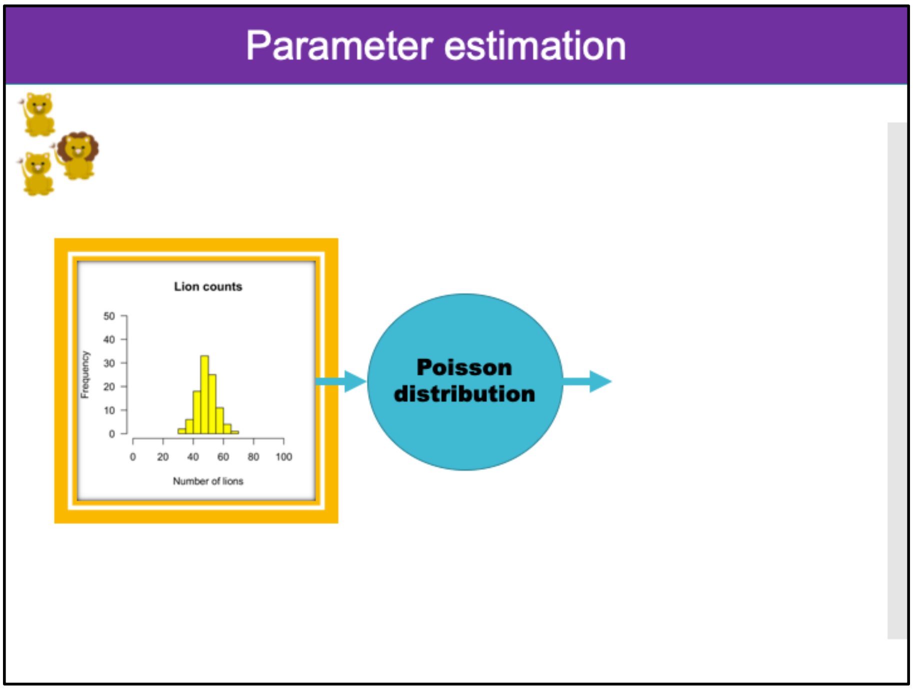
We want to estimate that parameter. We find the parameter which is most likely to give rise to our data



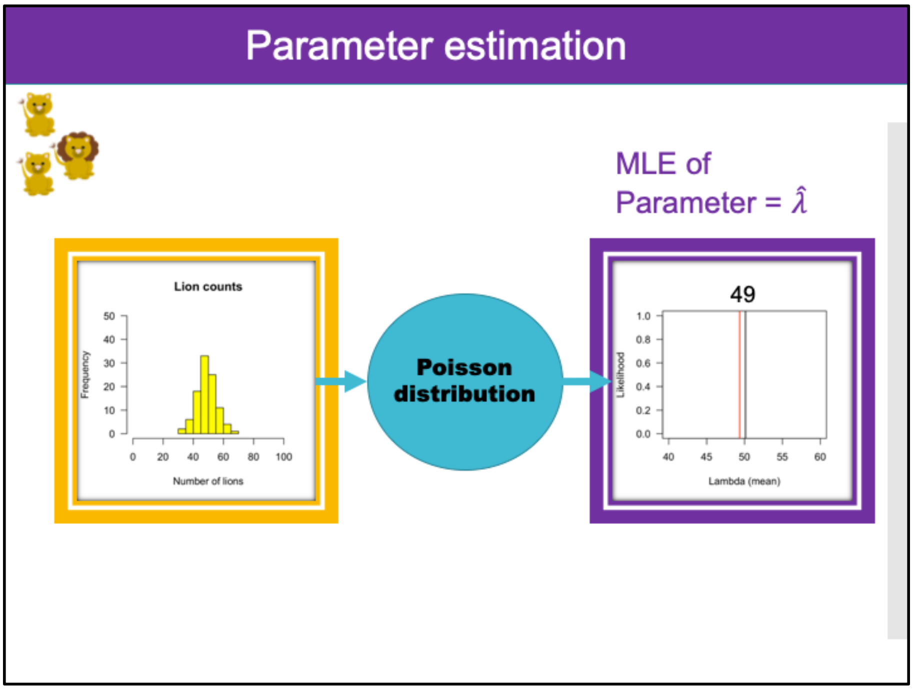
We want to estimate that parameter. We find the parameter which is most likely to give rise to our data



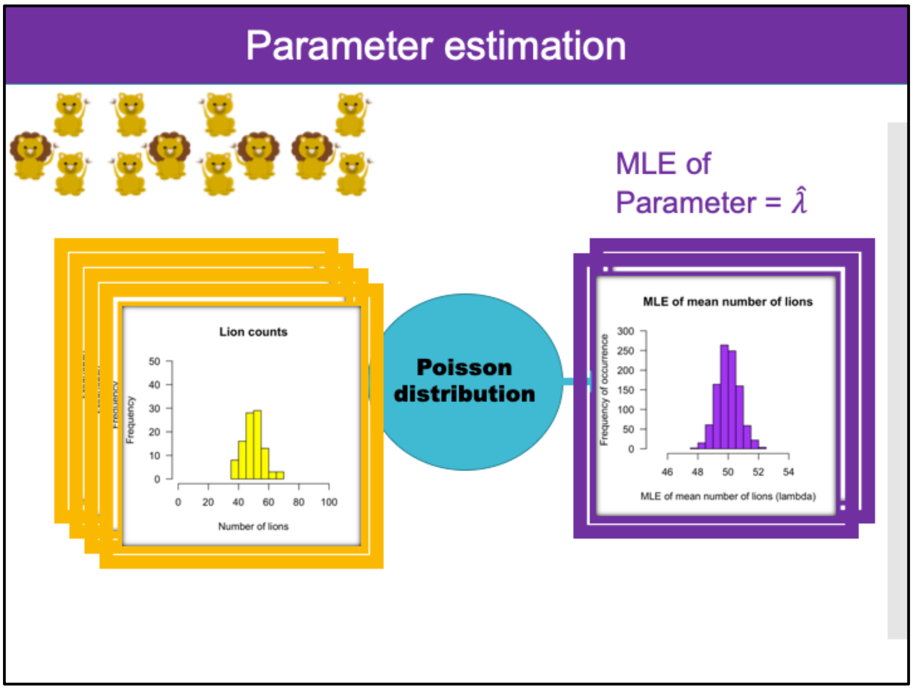
We use maximum likelihood estimation to get a maximum likelihood estimate of our parameter for this dataset



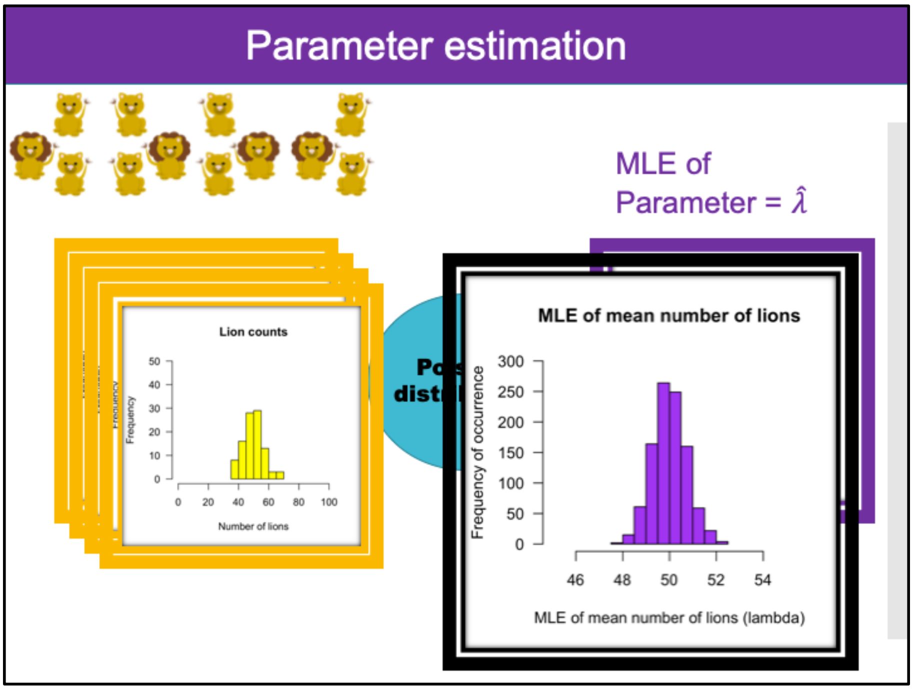
If you repeat this again with a different sample.



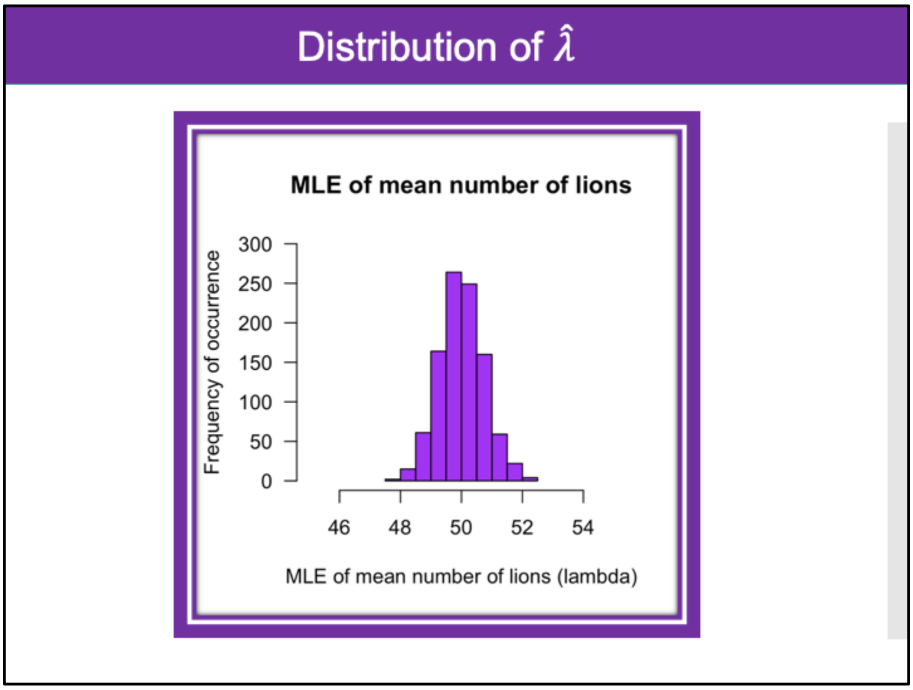
You will get a different estimate of the parameter in your model – because the likelihood is conditional on the data.



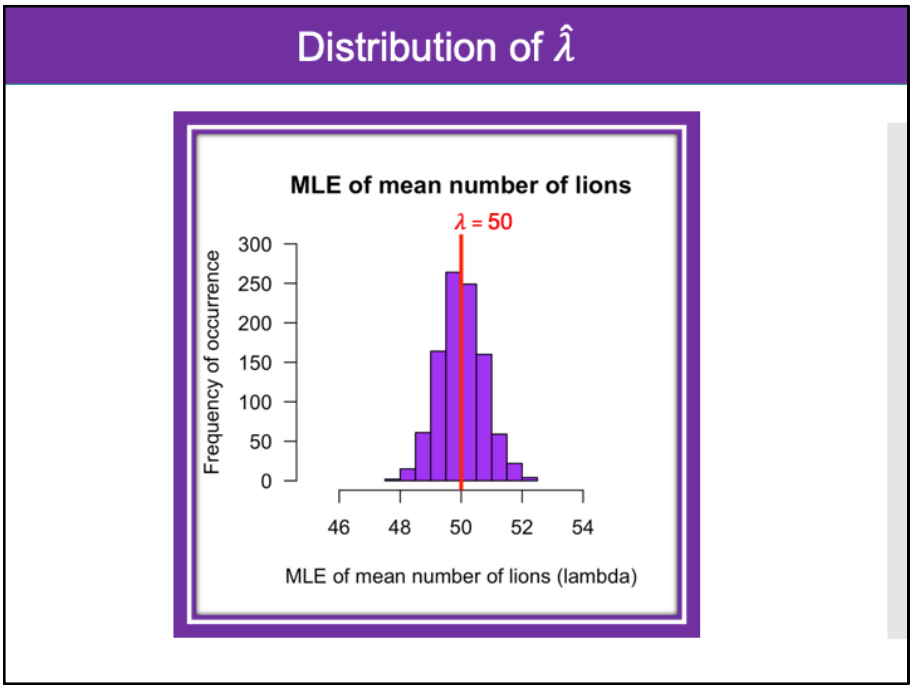
We use maximum likelihood estimation to get a maximum likelihood estimate of our parameter for this dataset



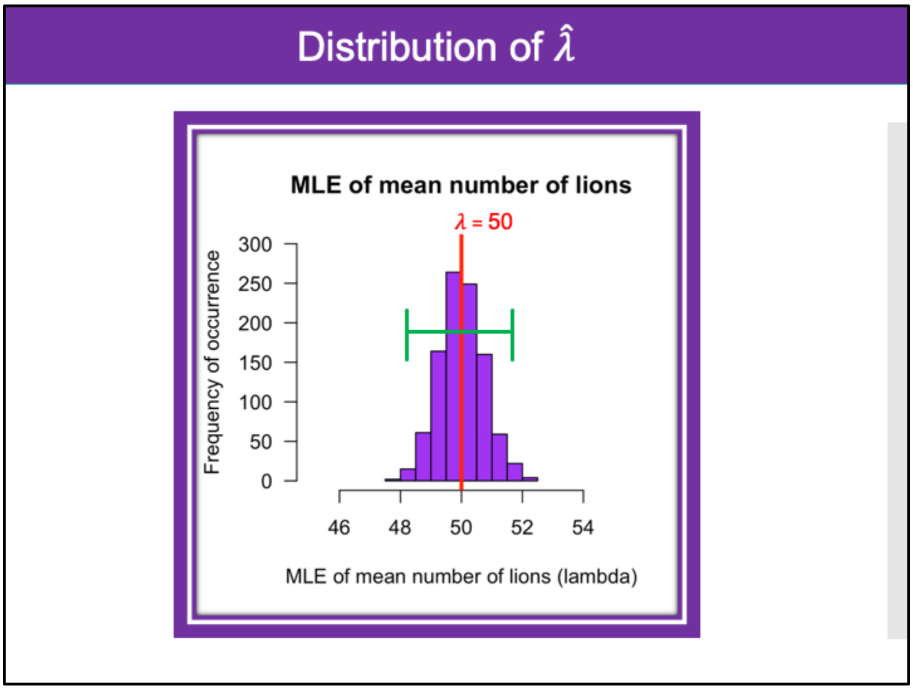
We use maximum likelihood estimation to get a maximum likelihood estimate of our parameter for this dataset



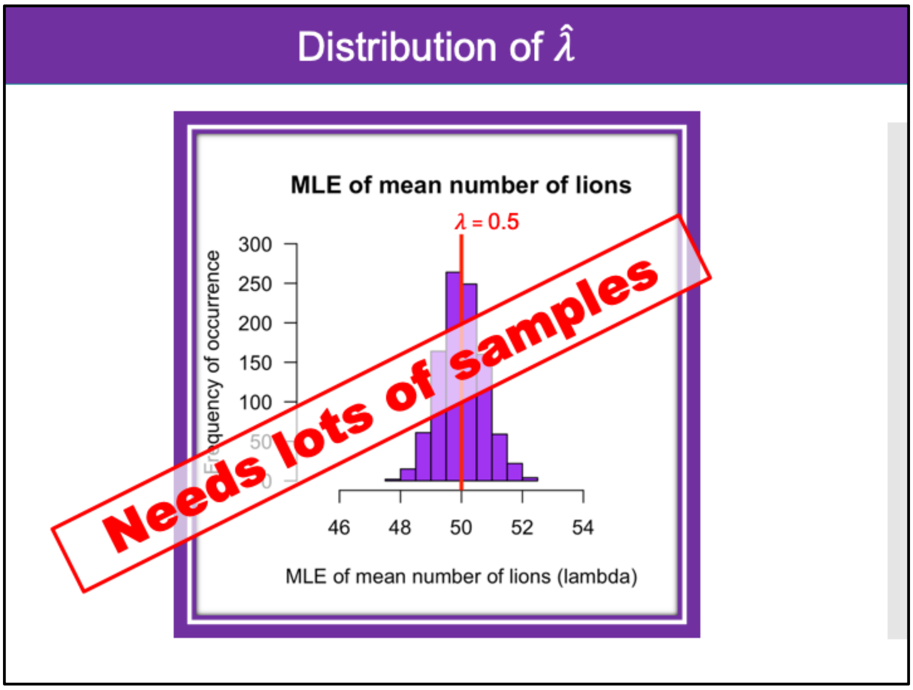
Here is the distribution of the estimates of the parameter. Can see out of 1000 samples, got around 50 about 260 times.



The mean is actually the true population level value of lambda



The mean is actually the true population level value of lambda



But to get this distribution requires a lot of simulation or many samples – this is rarely possible

Parameter estimation from a single sample

We really want a way to get that distribution of possible estimates from a single sample

Parameter estimation from a single sample

We need to:

- represent the distribution of the parameter mathematically
- and based on our single sample of data

We really want a way to get that distribution of possible estimates from a single sample

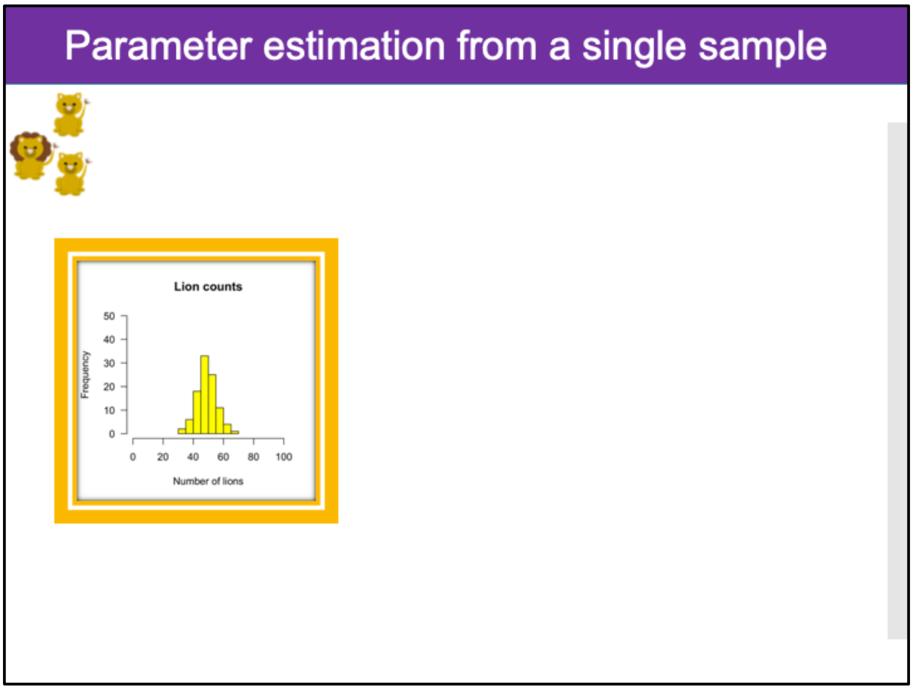
Parameter estimation from a single sample

We need to:

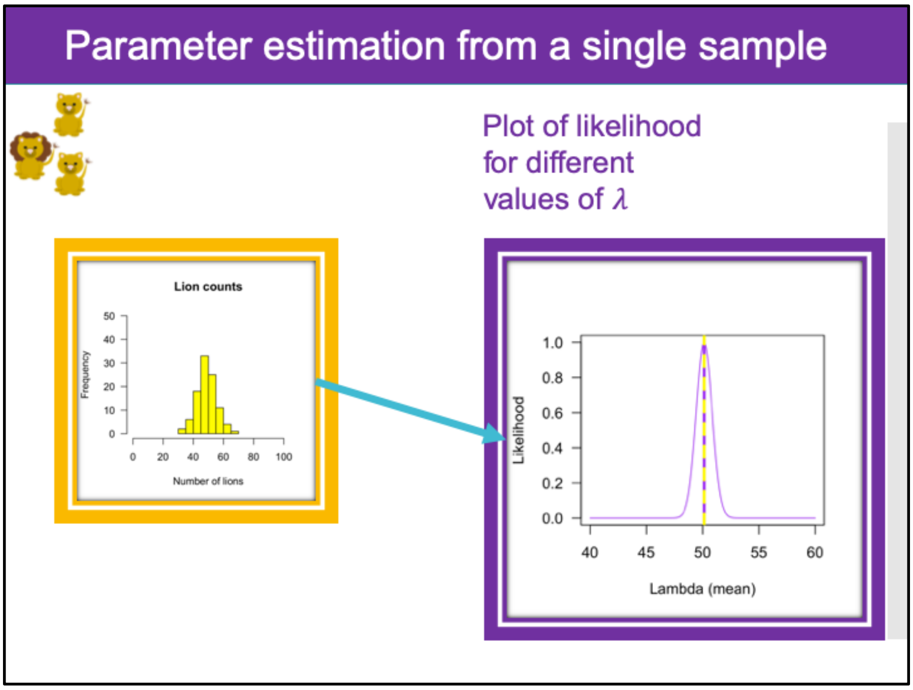
- represent the distribution of the parameter mathematically
- and based on our single sample of data

This is what we use the likelihood to do

This is where the likelihood and maximum likelihood estimation come in.



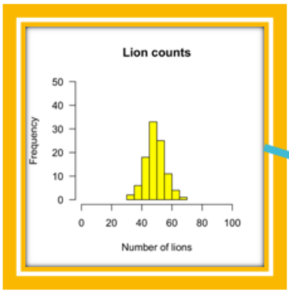
From a single sample

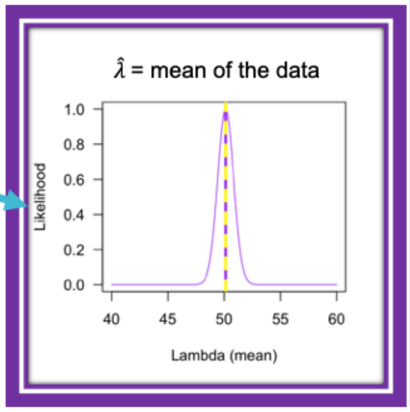


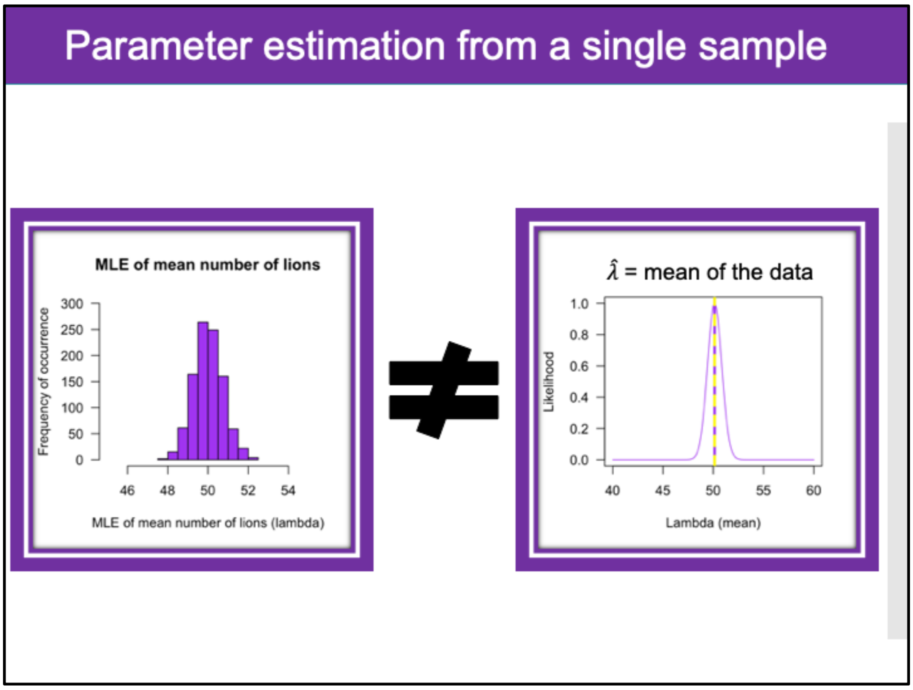
We produce a likelihood curve based on that data, we find the probability of getting this sample based on different values of the parameter and find the one that makes it most likely. Here that is also the mean of our sample. You can see that the estimate is closely tied to the data.

Parameter estimation from a single sample

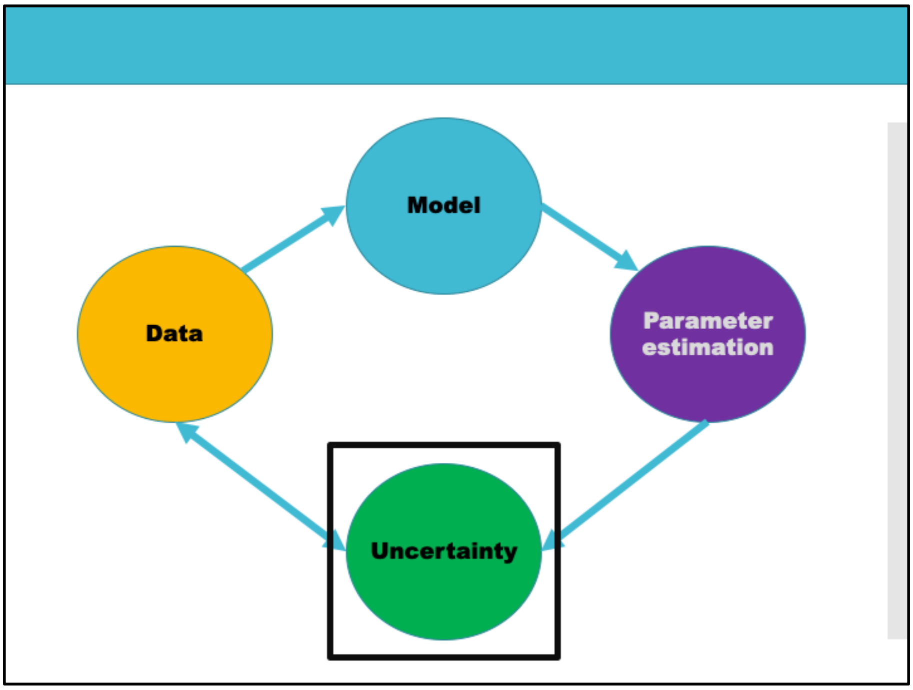
Plot of likelihood for different values of λ



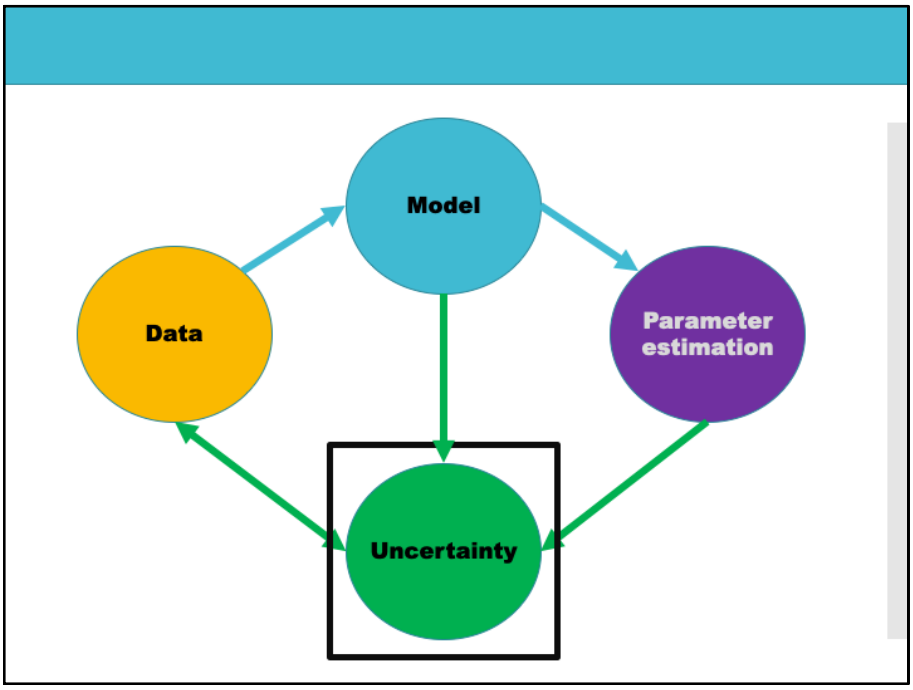




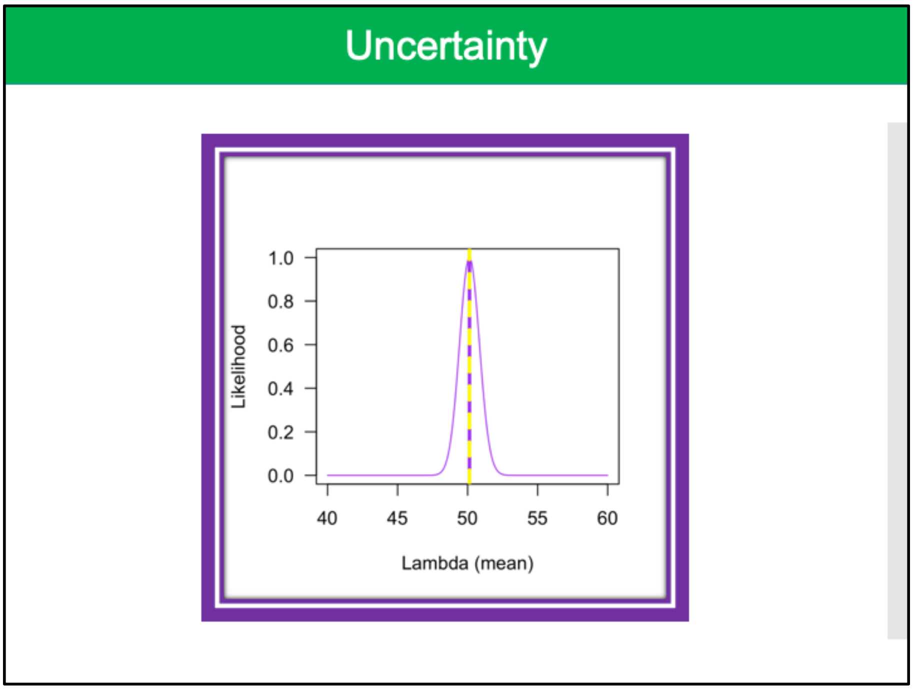
The actual distribution you would get by taking many samples is NOT the same as the one you get from maximum likelihood estimation. This is because when we estimate, it is relative to our data, that is the only information we have. So always centred on our MLE of the parameter – but tries to represent it from limited information



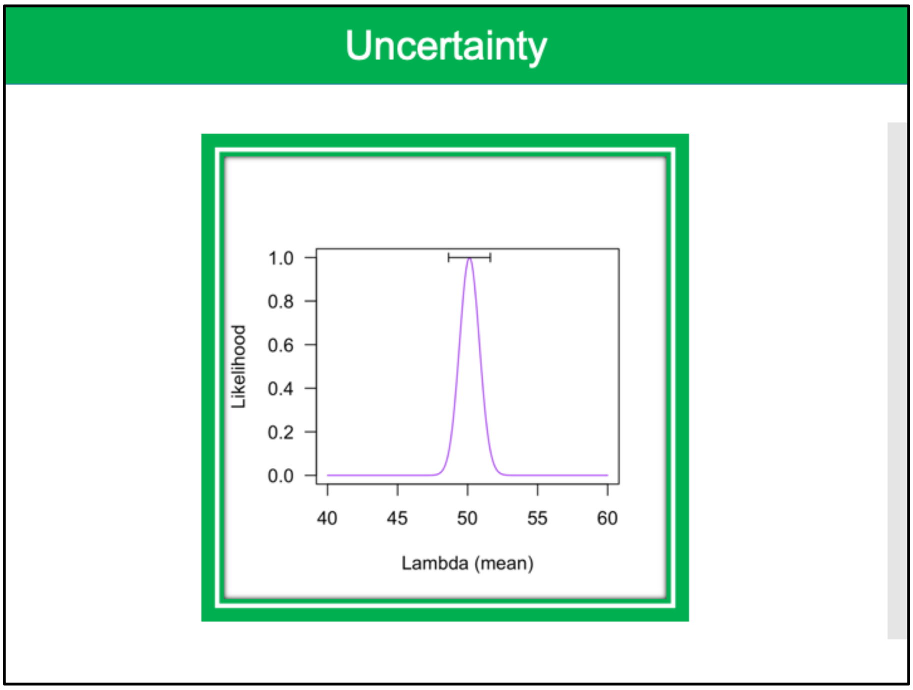
So, the final part to consider is uncertainty



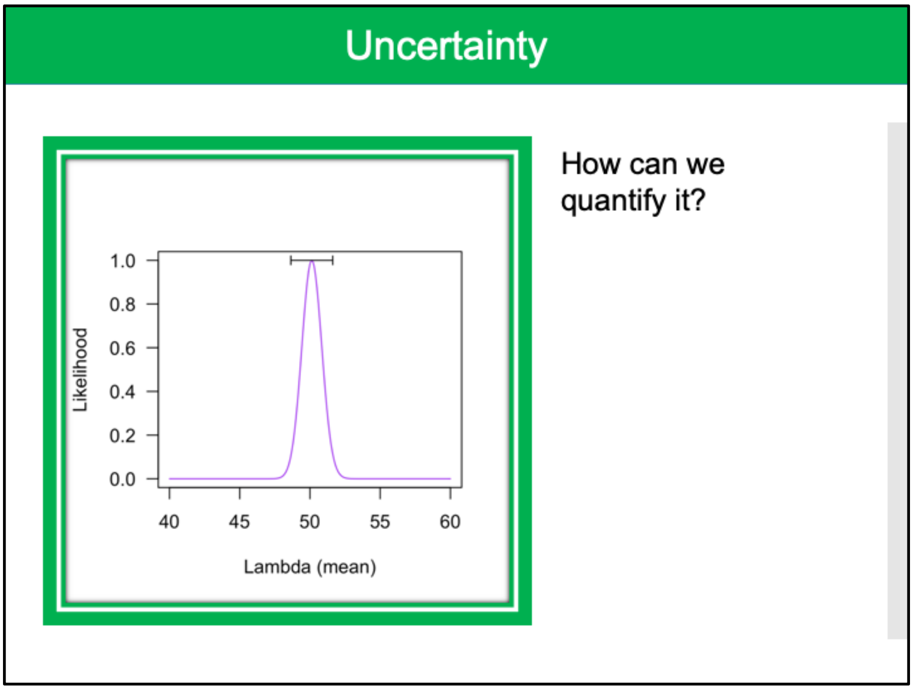
Actually it is influenced by everything else and is therefore really important!



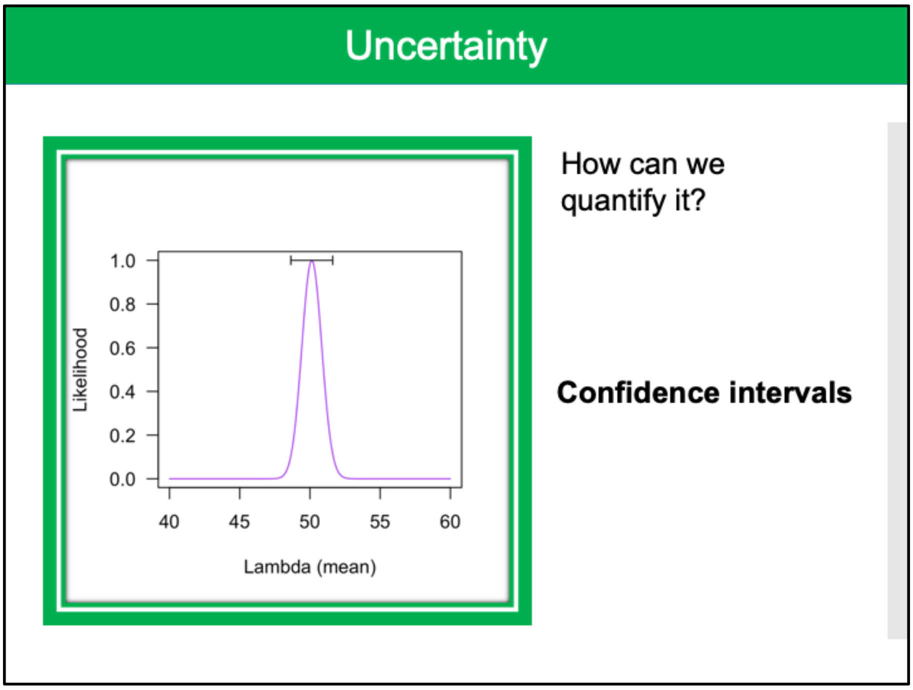
We are now back to our likelihood curve we can see there is another component to this distribution as well as the maximum



There is also a spread – this helps us to represent uncertainty in our estimate of the parameter

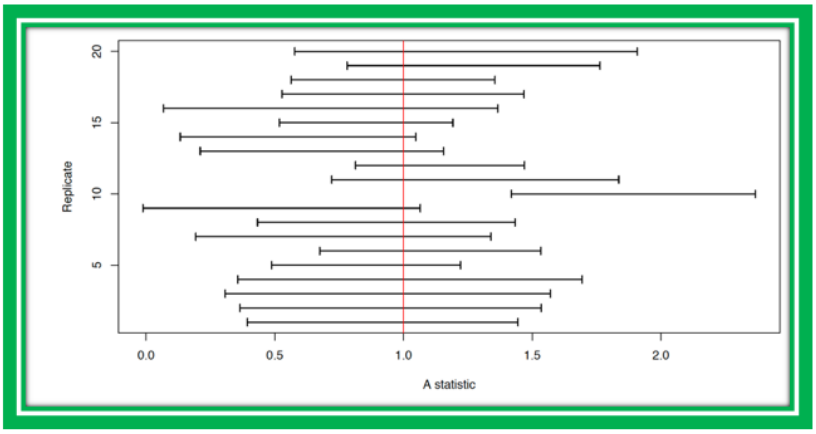


We represent it using confidence intervals



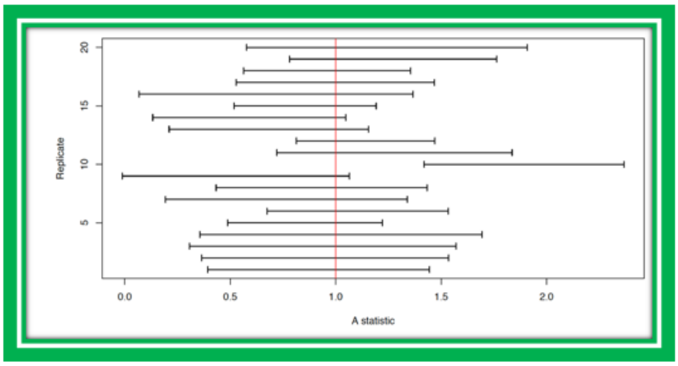
We represent it using confidence intervals – hopefully this is familiar from last week!

Confidence intervals



This was presented last week. Statistic on the x axis (here it would be our lambda) and repeated sample on the y axis. The horizontal bars are the confidence intervals and the red line is the true population statistic. Can see that not all confidence intervals include the truth.

Confidence intervals

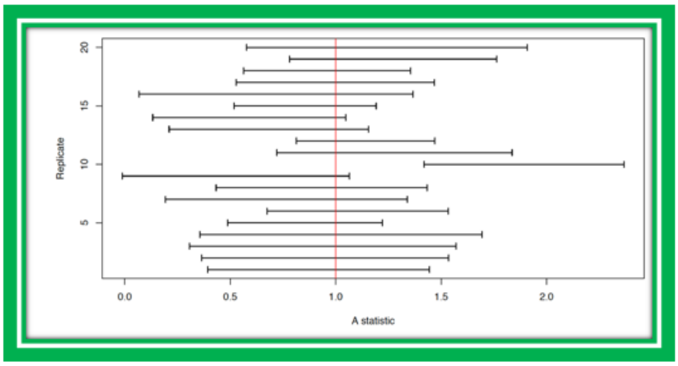


Confidence interval is used to indicate values for the true parameter that are more likely, given our data

IF you repeated your sampling many times and each time drew a confidence interval – 95% of the time (on average) the confidence interval would contain the true parameter value

Definition =

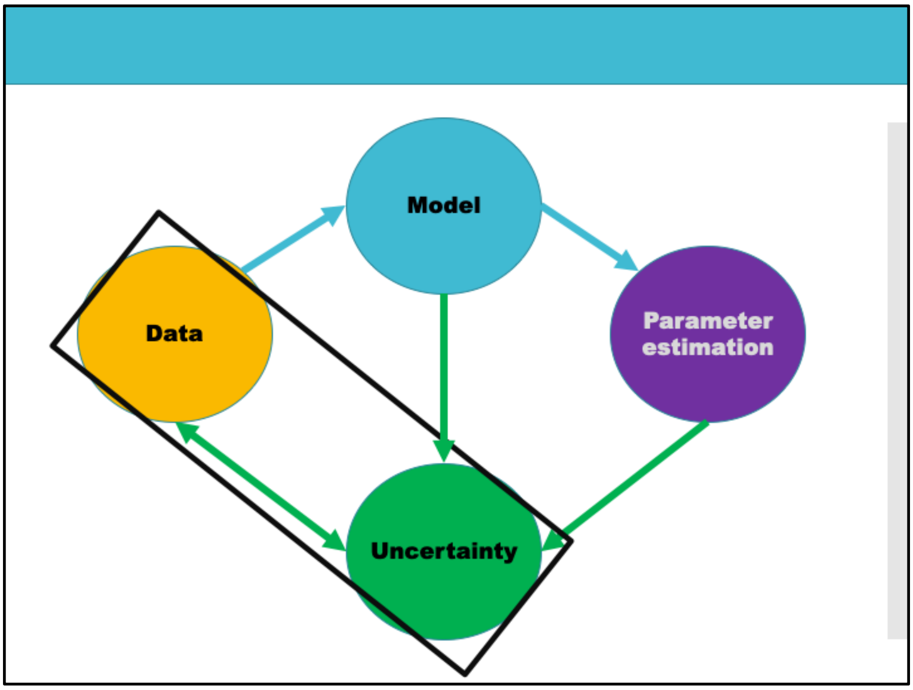
Confidence intervals



Confidence interval is used to indicate values for the true parameter that are more likely, given our data

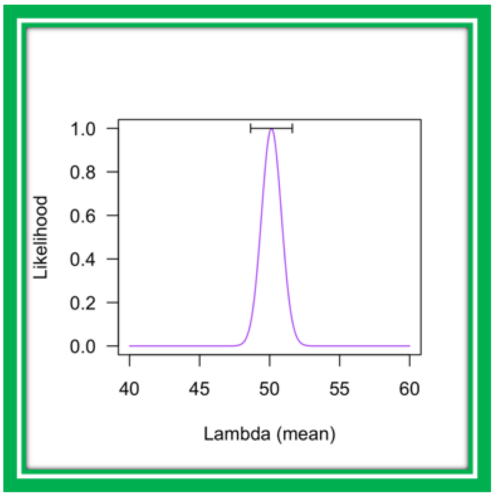
IF you repeated your sampling many times and each time drew a confidence interval – 95% of the time (on average) the confidence interval would contain the true parameter value **FREQUENTIST IDEA**

Definition = on slide



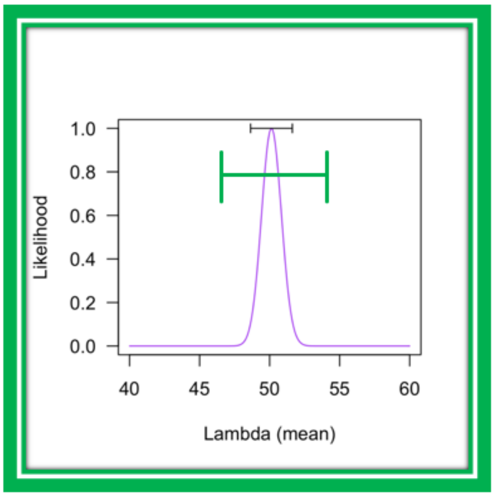
Finish off with the link between uncertainty and data

Uncertainty



This is the confidence interval for our data

Sample size and uncertainty



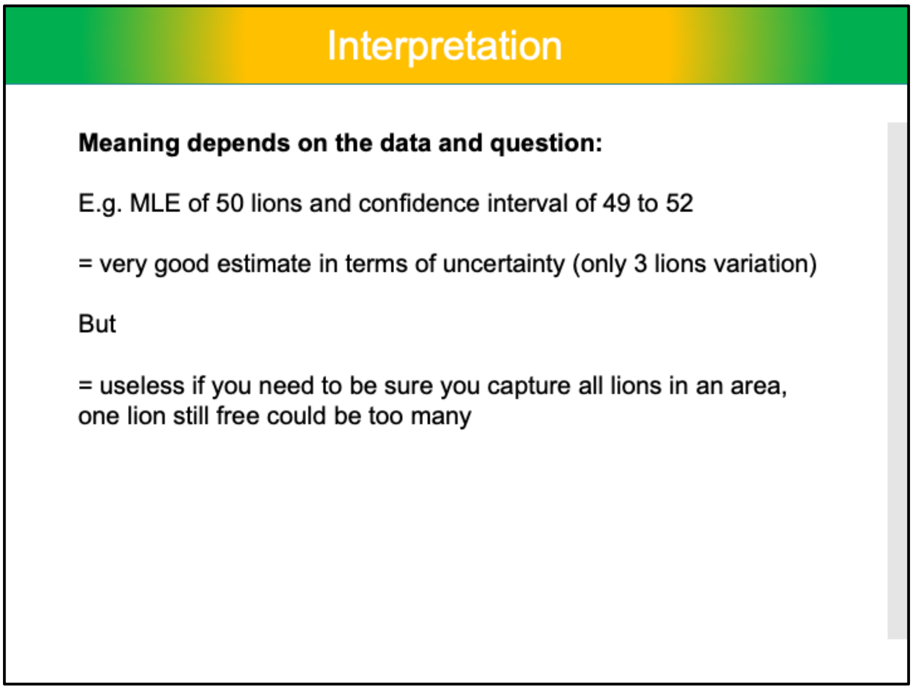
Less data = higher uncertainty in our estimates

This is the confidence interval for our data

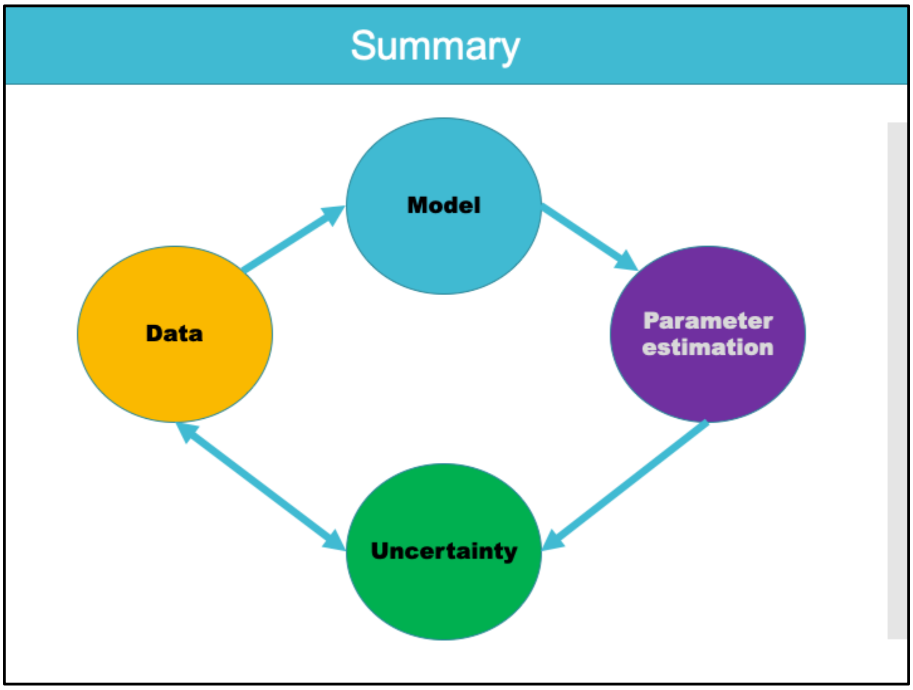
Interpretation

Meaning depends on the data and question

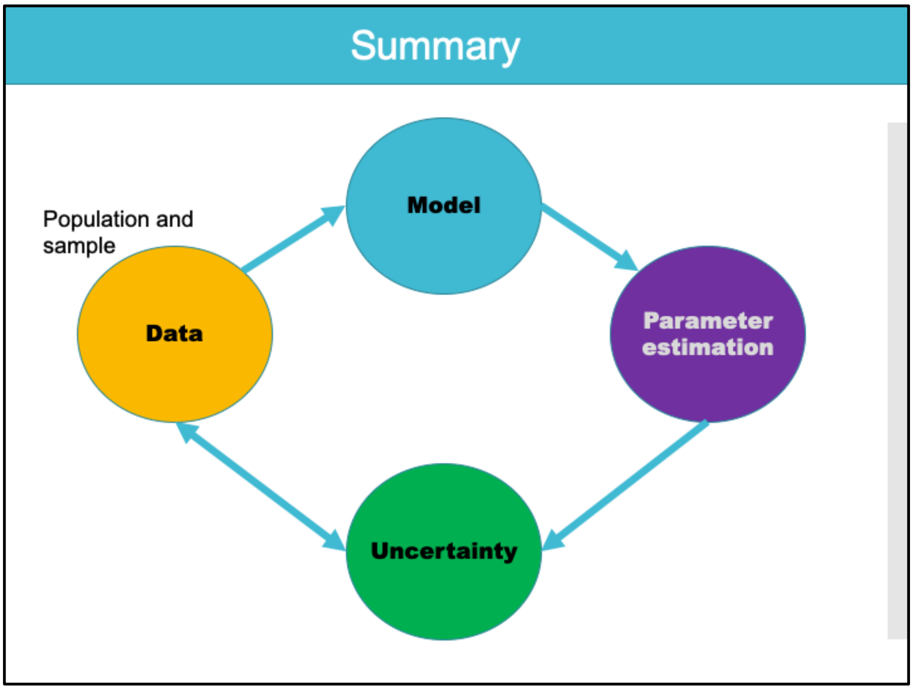
How you interpret the results of maximum likelihood parameter estimation depends on your data and question



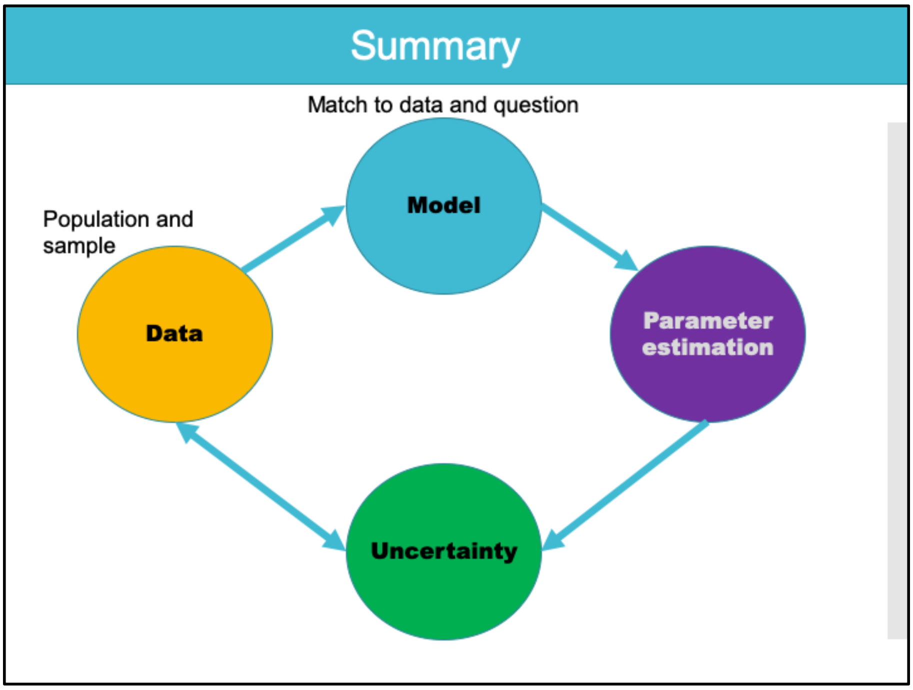
How you interpret the results of maximum likelihood parameter estimation depends on your data and question



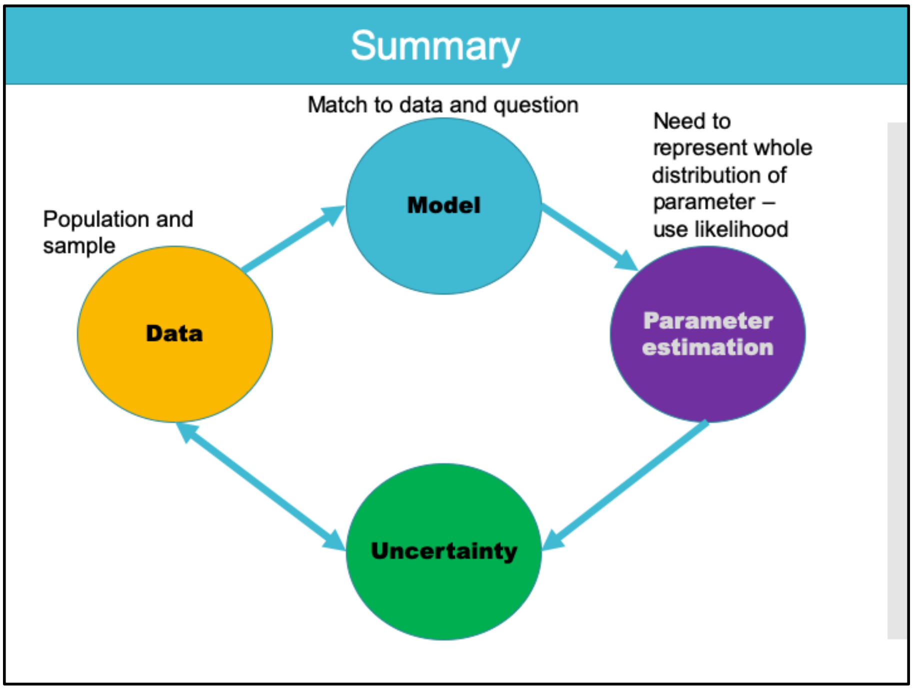
Summarise the lecture on the schematic



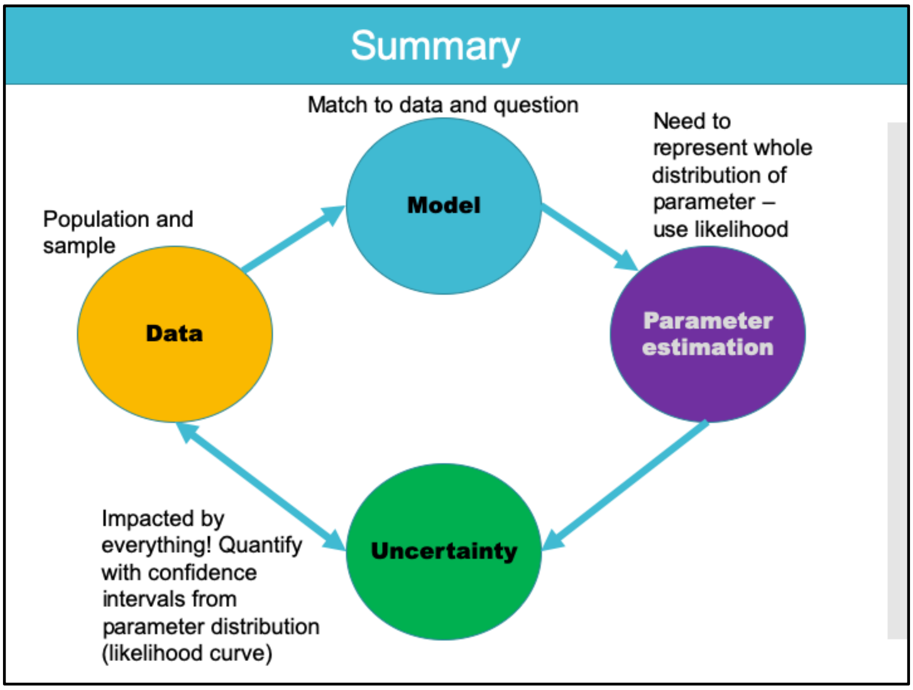
Summarise the lecture on the schematic



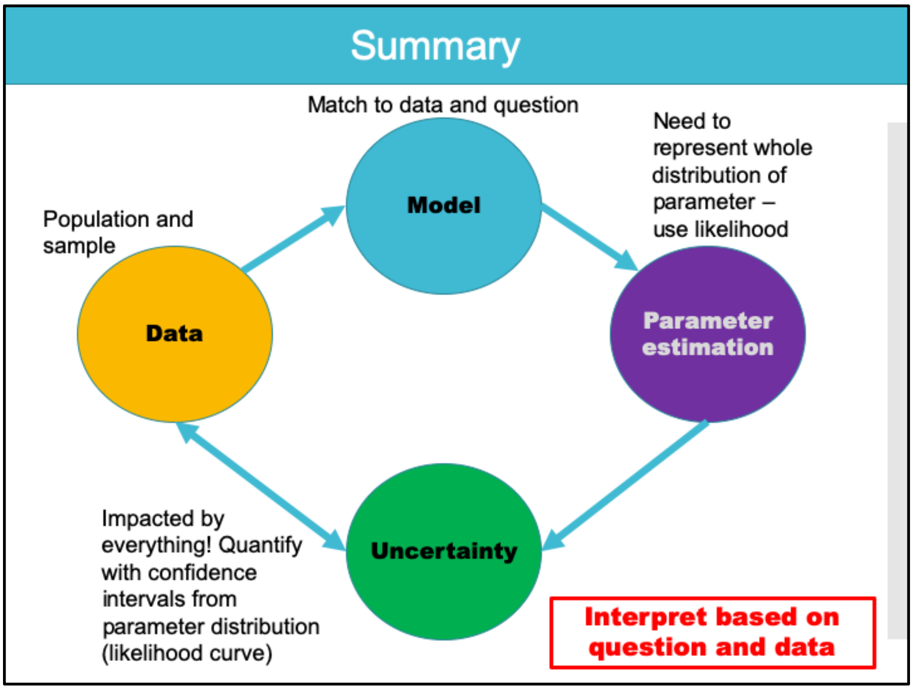
Summarise the lecture on the schematic



Summarise the lecture on the schematic



Summarise the lecture on the schematic



Summarise the lecture on the schematic

EXERCISE TASK

On your table write in your own words:

“What maximum likelihood estimation is and why we use it in statistical modelling”

ASK any questions!!!