

# An introduction to parameter estimation in outbreak models (Fitting models to data)

MATHEMATICAL MODELING FOR INFECTIOUS DISEASE  
PLANNING IN AFRICA  
(60 minutes)

# Learning objectives

At the end of this lecture, you will:

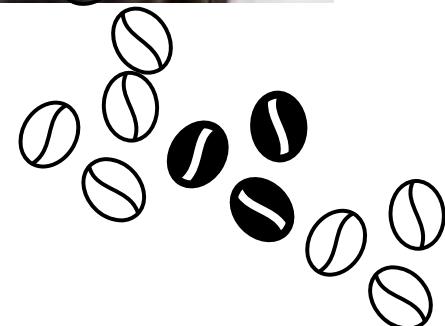
1. Understand the **need** for parameter estimation in outbreak models
2. Identify the **contexts** in which parameter estimation is needed
3. Understand the **basic idea** underlying parameter estimation methods
4. Be introduced to **examples** of parameter estimation in outbreak models

# Outline

1. Intuition
2. Introduction to parameter estimation methods
3. Parameter estimation examples with Ebola and hepatitis A

# Intuition

# Who spilled the beans?



Cat A

or

Cat B

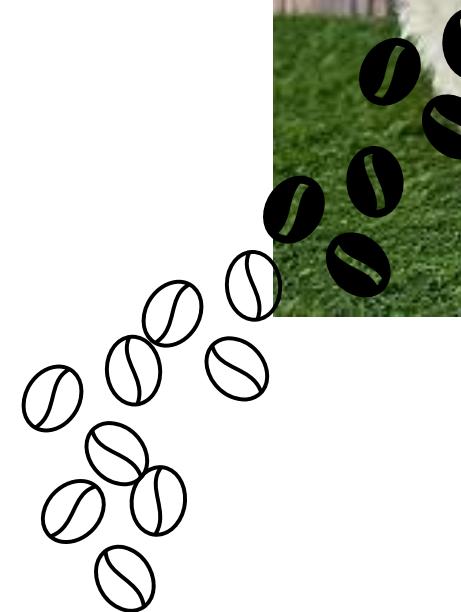


# Who spilled the beans?



Dog A

or



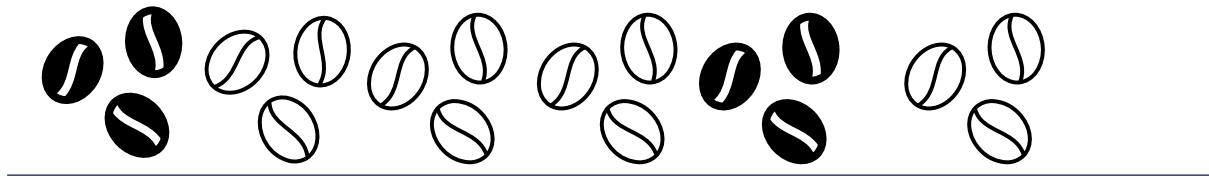
Dog B

# Who spilled the beans?

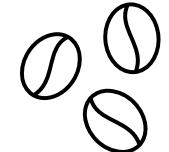


Parameter

Estimation

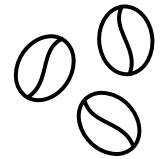


# Bean spilling and data generation



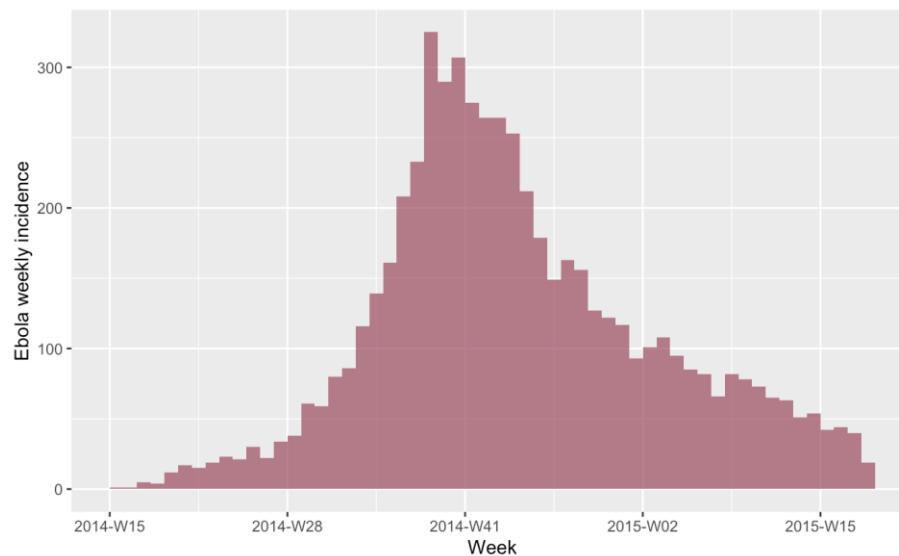
- Bean-spilling process is analogous to the data-generation process in the real world.
  - Examples of such data: incidence, seroprevalence,...
- Mechanisms which lead to the observed data are often **not observed**.
- To analyze these mechanisms, we often design **models** which seek to approximate the underlying data-generation process.
- **Models** are often characterized by **parameters**.
- *True parameter values are typically unknown since entire disease transmission process process is typically unobserved.*

# The problem

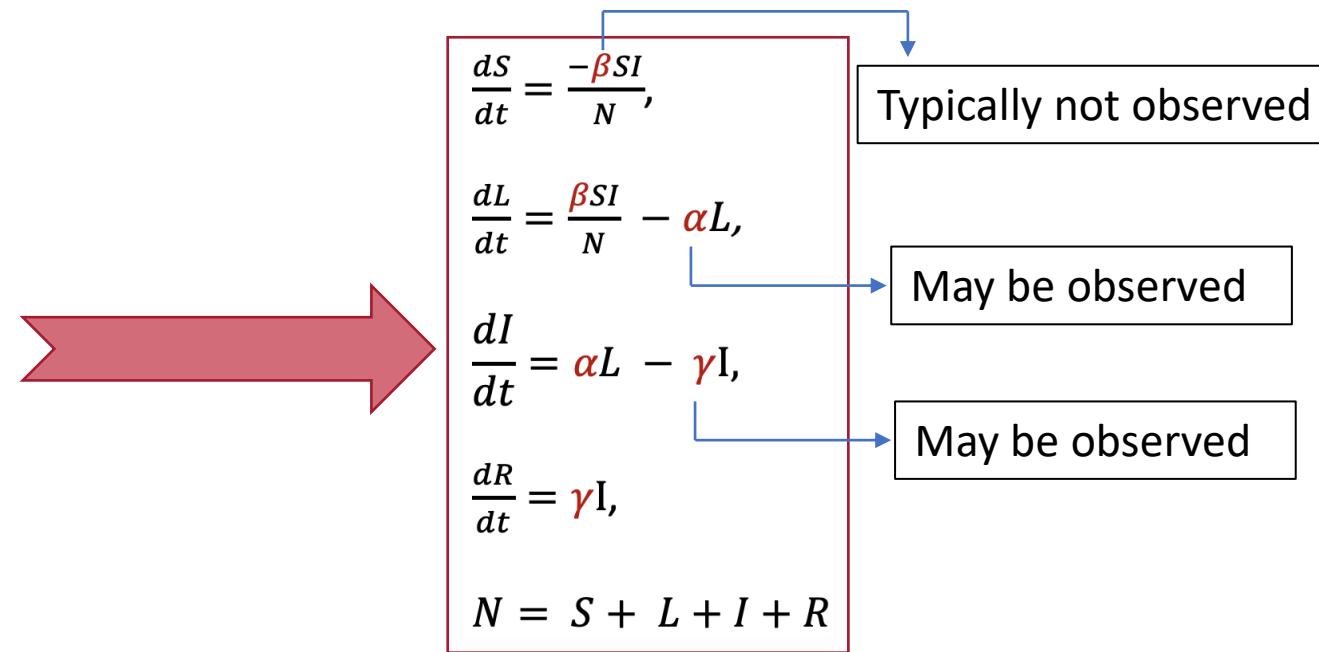


True parameter values are typically unknown since entire disease transmission process is not observed.

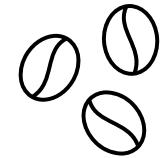
Step 1: Observe data ( $D$ )



Step 2: Design model ( $M$ )



# The problem



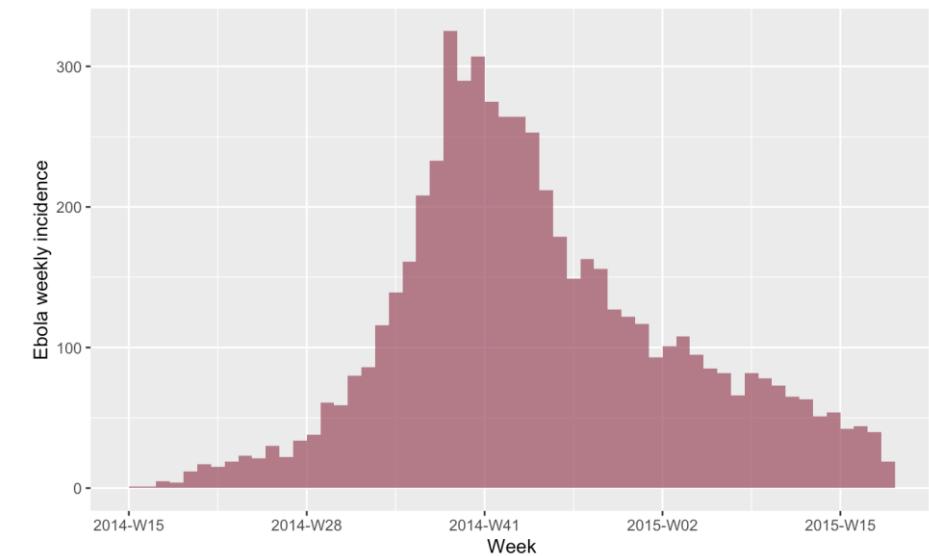
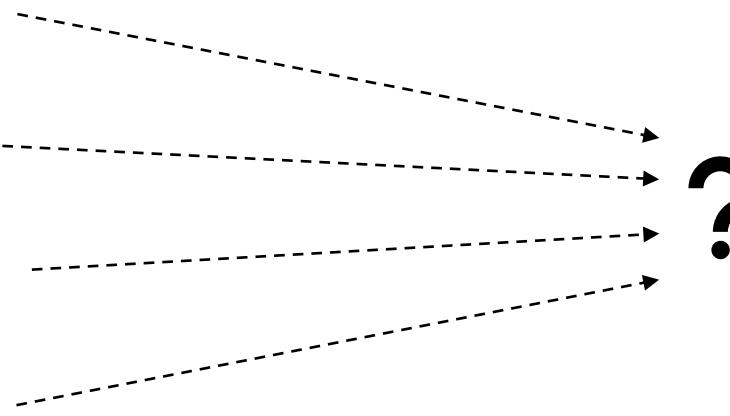
True parameter values are typically unknown since entire disease transmission process is not observed.

*Step 3: Determine “who” spilled the data*

$$\beta = 2, \gamma = 3, \alpha = 0.4$$

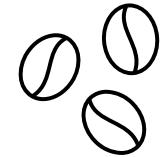
$$\beta = 0.2, \gamma = 3, \alpha = 0.07$$

$$\beta = 5, \gamma = 13, \alpha = 8$$



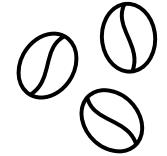
Which combination of values gave rise to the data?

# The problem



- Which combination of values gave rise to the data?
  - We do not know for certain, but we can estimate.
- The parameter estimation problem: Given some observed data  $D$ , and a model  $M_\theta$  parameterized by  $\theta$ , estimate the values of  $\theta$  for which  $M_\theta$  best describes  $D$  to a given confidence level.
- In statistics, *estimation* is making inferences: statements about the true value/distribution of a parameter based on a sample from the population.

# Revisit pre-session video on parameter estimation



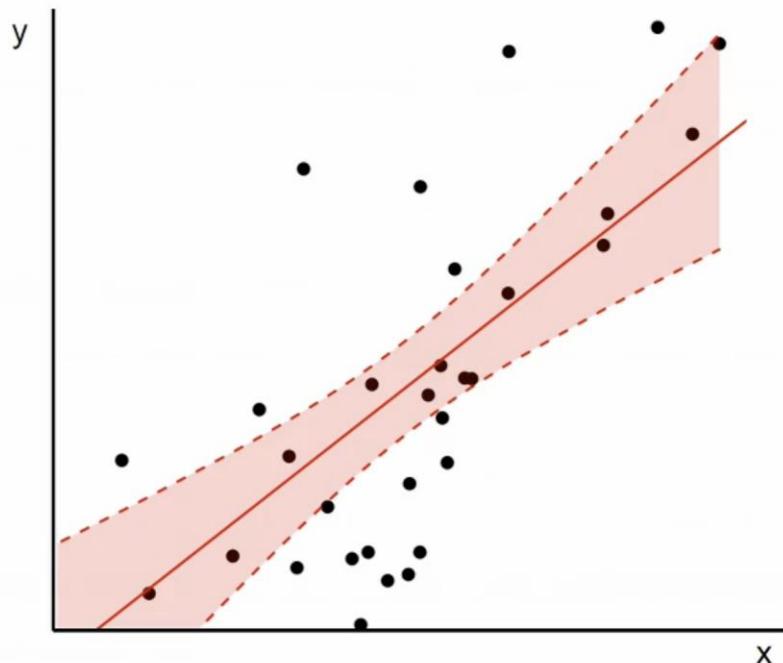
<https://ccdd.hsph.harvard.edu/introduction-to-infectious-disease-modeling/week-5-model-fitting-and-parameter-estimation/>

# Parameter estimation methods

# Basic idea

## Simple example: Linear regression

---



6

Define the model:  
 $y = \alpha + \beta x + \epsilon$

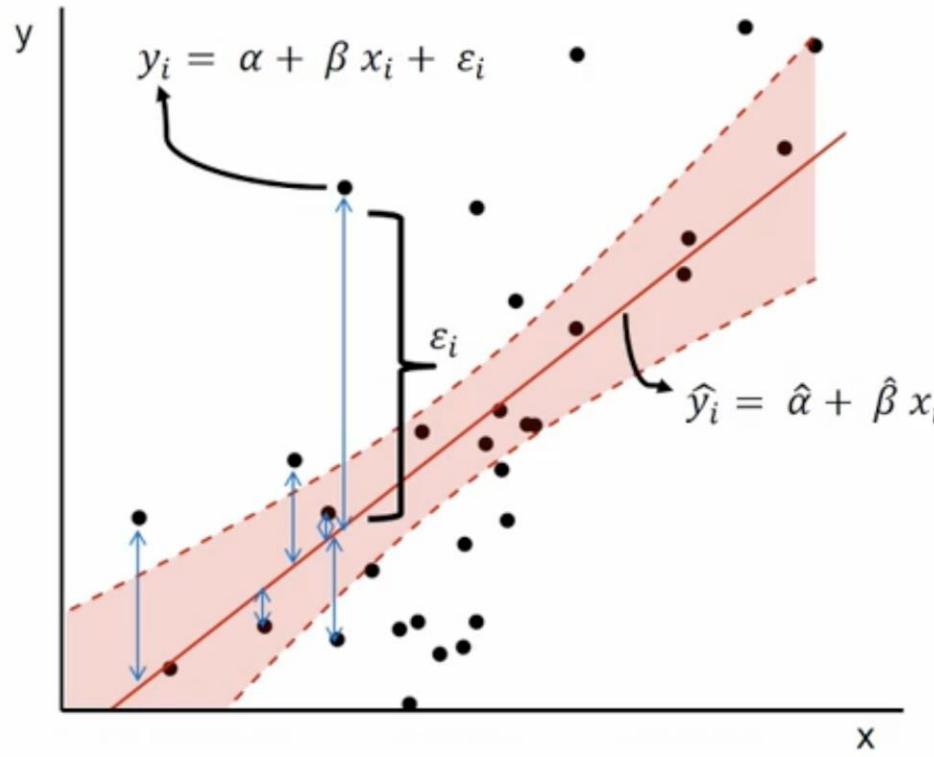
Estimate parameters:  
 $\alpha$  = intercept  
 $\beta$  = slope

Evaluate model fit and  
quantify uncertainty:  
(for e.g. 95% confidence  
intervals)

# Basic idea

## Fitting to data: Least-Squares method

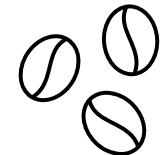
---



Minimize the sum of squared residuals

$$\sum_i^n (y_i - \hat{y}_i)^2 = \sum_i^n (\varepsilon_i)^2$$

# Two schools of thought



- *Frequentist* paradigm
  - Parameter,  $\theta$ , is treated as a fixed value
  - “*To learn about  $\theta$ , the data (and model) are all I have*”.
- *Bayesian* paradigm
  - Parameter,  $\theta$ , is treated as a random variable that has a probability distribution
  - “*To learn about  $\theta$ , my prior beliefs about  $\theta$ , the data (and model) are all I have*”.

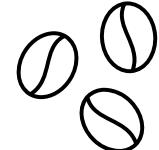
# Some parameter estimation methods

Bayesian methods	Frequentist methods
Markov Chain Monte Carlo - Random walk - Gibbs sampler - Metropolis-Hastings algorithm - Hamiltonian Monte Carlo...	Maximum likelihood estimation
Sequential Monte Carlo (SMC)	z-, t-, and F- tests (Frequentist hypothesis testing)
Approximate Bayesian Computation (ABC)	

- Many of these algorithms are implemented in packages in R.
- Important to understand the theory so results are interpreted correctly

# Example 1: Ebola virus transmission

# Estimating the values of $\beta$ for Ebola virus transmission



- What do we have?

Observed data ( $D$ )

time <int>	incidence <dbl>
30	7
31	8
32	9
33	11
34	12
35	14
36	15
37	17
38	20
39	22

Previous 1 2 3 4 5 6 ... 37 Next

+ SLIR model ( $M$ ) + Values for  $\gamma$  and  $\alpha$

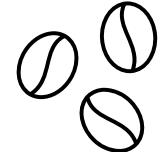
$$\begin{aligned}\frac{dS}{dt} &= -\beta SI, \\ \frac{dL}{dt} &= \frac{\beta SI}{N} - \alpha L, \\ \frac{dI}{dt} &= \alpha L - \gamma I, \\ \frac{dR}{dt} &= \gamma I, \\ N &= S + L + I + R\end{aligned}$$

+

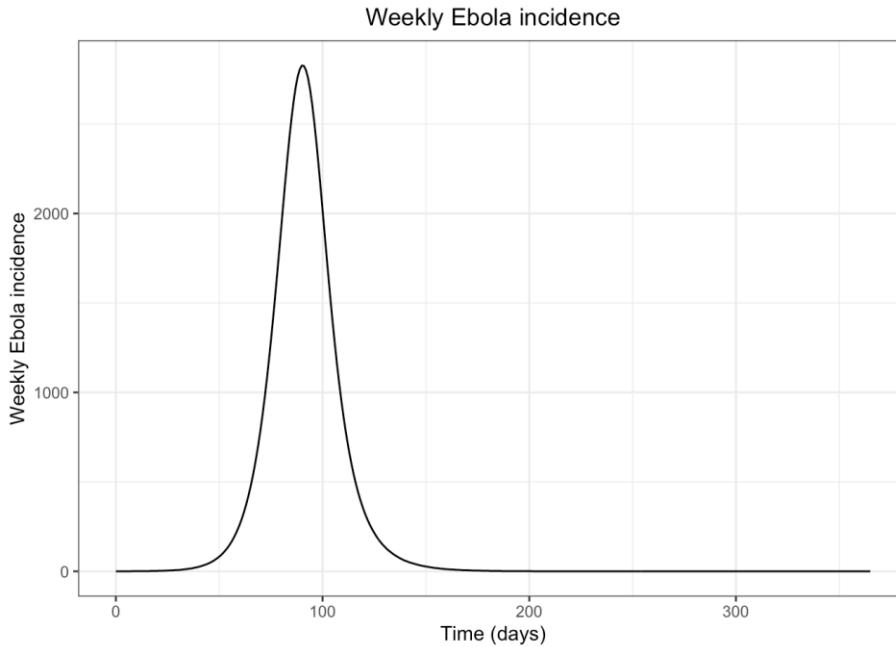
$$+ \quad \alpha = \frac{1}{3}, \gamma = \frac{1}{10}$$

- We need to estimate the value of  $\beta$

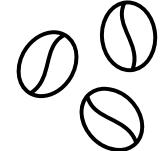
# Estimating the value of $\beta$ for Ebola virus transmission in R



## 1. Explore the data



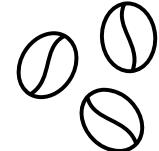
# Estimating the value of $\beta$ for Ebola virus transmission in R



1. Explore the data
2. Write and code your model function

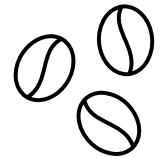
```
SLIR_model <- function(t, x, parms, alpha = 1/3, gamma = 1/10) {  
  with(as.list(c(parms, x)), {  
  
    N <- S + L + I + R  
    dS <- - (beta*S*I) / N  
    dL <- (beta*S*I) / N - alpha*L  
    dI <- alpha*L - gamma*I  
    dR <- gamma*I  
  
    states <- c(dS, dL, dI, dR)  
  
    return(list(states))  
  })  
}
```

# Estimating the value of $\beta$ for Ebola virus transmission in R

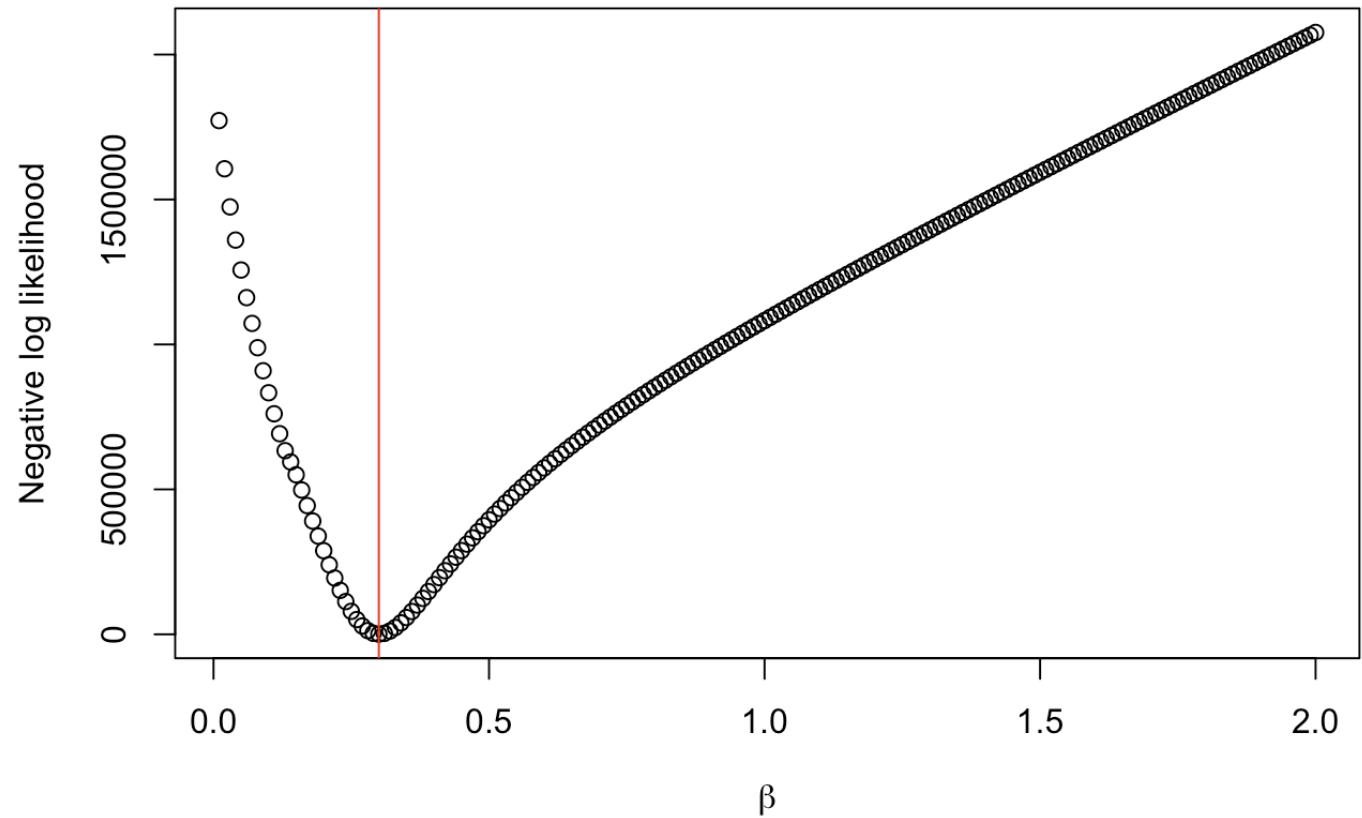


1. Explore the data
2. Write and code your model function
3. Choose a parameter estimation method. [Here, we choose the *maximum likelihood* method and hence we define a likelihood function]

# Estimating the value of $\beta$ for Ebola virus transmission in R

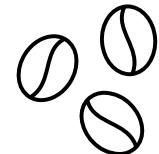


1. Explore the data
2. Write and code your model function
3. Choose a parameter estimation method. [Here, we choose the *maximum likelihood* method and hence we define a likelihood function]
4. Perform parameter estimation to find the “best” value of  $\beta$ .

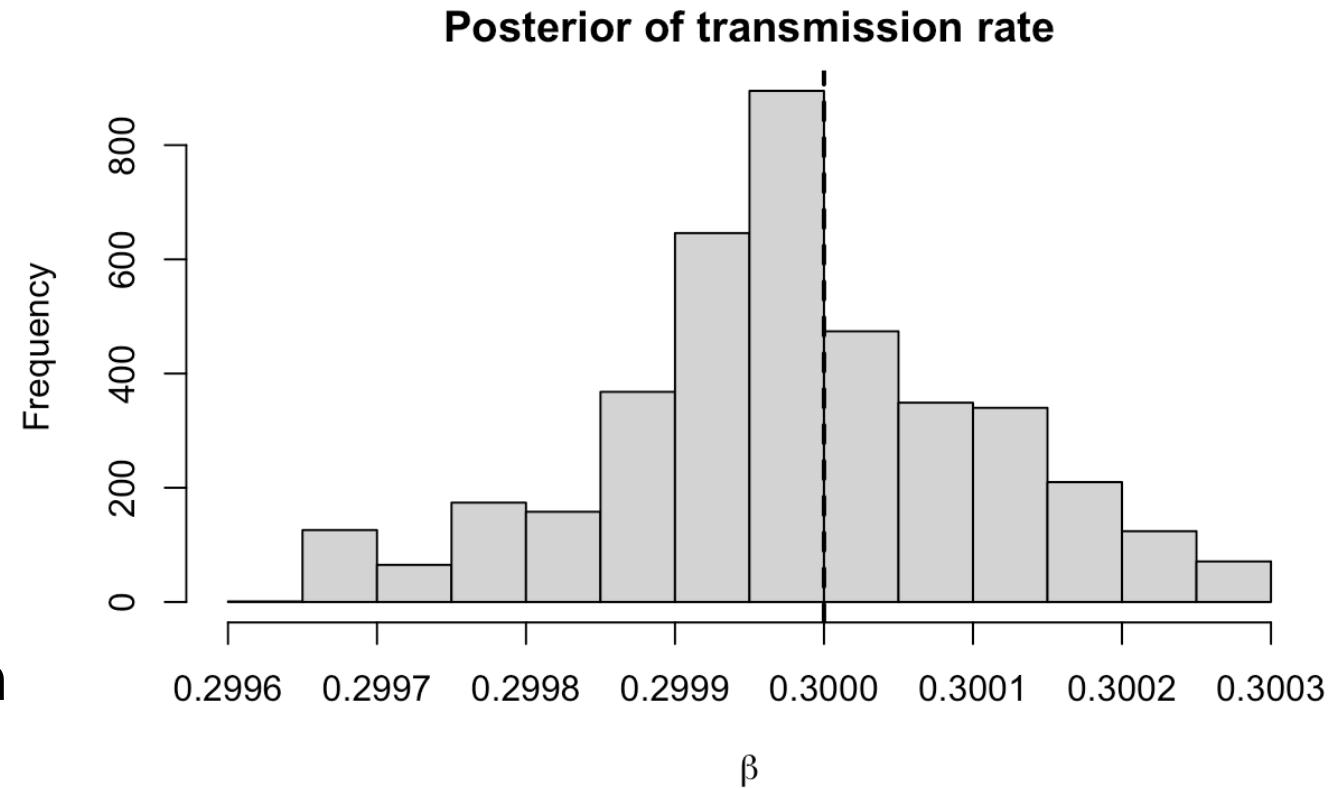


$$\beta_{MLE} \approx 0.3$$

# Estimating the value of $\beta$ for Ebola virus transmission in R

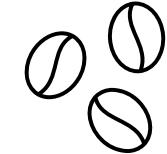


1. Explore the data
2. Write and code your model function
3. Choose a parameter estimation method. [Here, we choose the **MCMC** method]
4. Perform parameter estimation to find the “best” **distribution** of  $\beta$ .

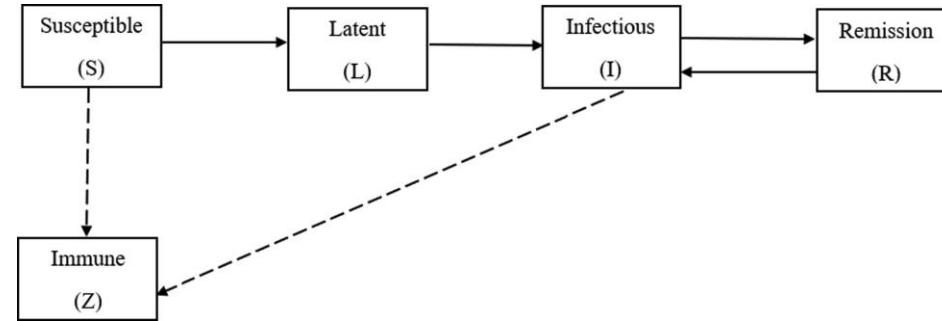
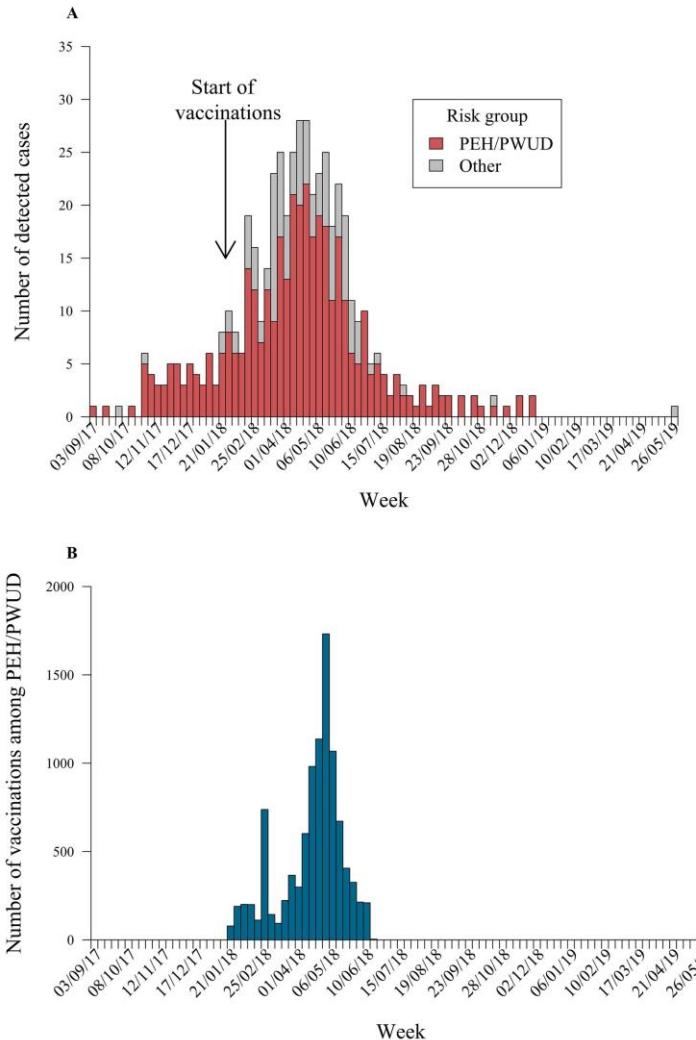


## Example 2: Hepatitis A virus (HAV) transmission

# Resources for parameter estimation

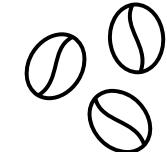


Observed data ( $D$ ) + SLIRZ model for HAV ( $M$ ) + Some parameter values



Duration of latent period (in weeks)	$1/\alpha$	1.57	[19]
Duration of infectious period (in weeks)	$1/\gamma$	4.64	[19]
Duration of remission period (in weeks)	$1/\sigma$	4.30	[20]
Probability of experiencing a relapse	$1-\eta$	0.11	[20]
Proportion of initially immune individuals	$\varepsilon$	0.43	[9], [10], [11]
First dose vaccine efficacy (%)	$\tau$	90	See Section 3
Fraction of vaccine doses given to at-risk individuals	$\omega$	1	Assumed
PEH/PWUD population size	$N$	69,862	[24], [26], [25], [27]

# Estimating $\beta(t)$ and $I(0)$



1. Explore the data

2. Write and code your model function

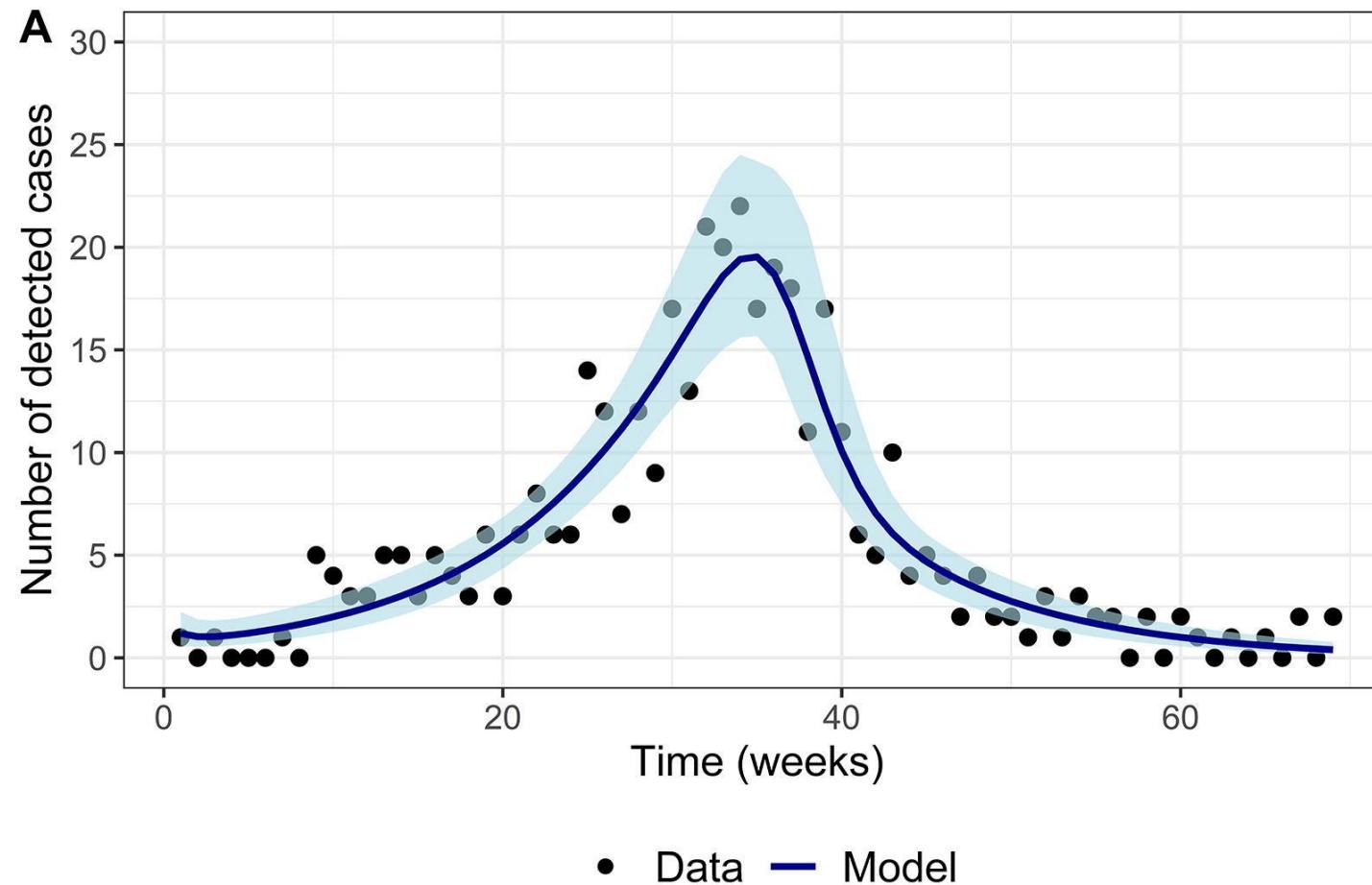
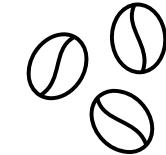
3. Choose a parameter estimation method. [Here, we choose the *maximum likelihood* method and hence we define a likelihood function]

4. Perform parameter estimation to find the “best” parameter values.

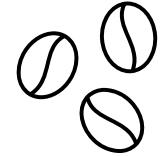
Table 1. Descriptions, values and references for model parameters. For estimated parameters, values are estimates, with 95% confidence intervals in parentheses.

Description	Symbol	Value	Reference
Effective contact rate at the start of the outbreak	$\beta_s$	0.61 (0.54– 0.67)	Estimated
Effective contact rate later in the outbreak	$\beta_l$	0.12 (0.05– 0.18)	Estimated
Transition midpoint time from $\beta_s$ to $\beta_l$ (in weeks)	$t^*$	36 (33–39)	Estimated
Speed of transition from $\beta_s$ to $\beta_l$	$c$	0.75 (0.24– $\infty$ )	Estimated
Number of infectious individuals in week 0	$I_1(0)$	0.92 (0.63– 2.01)	Estimated

# Model fit given estimated values

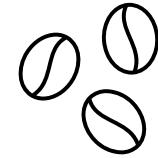


# Later...



- Quantifying uncertainty in parameter values
- Assessing sensitivity of models to changes in parameter values

# Resources



- **Parameter estimation in epidemic models**
  - “Introduction to inference: parameter estimation” course by Aaron King. <https://kingaa.github.io/short-course/parest/parest.html>
  - A general primer: Gerardo Chowell (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, Volume 2, Issue 3, Pages 379-398. <https://doi.org/10.1016/j.idm.2017.08.001>
- **Markov Chain Monte Carlo**
  - A gentle introductory article - Jason Brownlee: <https://machinelearningmastery.com/markov-chain-monte-carlo-for-probability/>
  - A good textbook with code examples - Kruschke (2015). Doing Bayesian data analysis. A Tutorial with R, JAGS, and Stan. [https://nyu-cdsc.github.io/learningr/assets/kruschke\\_bayesian\\_in\\_R.pdf](https://nyu-cdsc.github.io/learningr/assets/kruschke_bayesian_in_R.pdf)
- **Bayesian vs. Frequentist**
  - Cassie Kozrykov: <https://towardsdatascience.com/statistics-are-you-bayesian-or-frequentist-4943f953f21b>
  - Jake VanderPlas: <https://conference.scipy.org/proceedings/scipy2014/vanderplas.html>