# Parameter Estimation for a Single-Outbreak Model

Ariel Cintrón, Ph.D.

## Single Outbreak Susceptible-Infective-Recovered (SIR) Model

$$\frac{dS}{dt} = -\beta S \frac{I}{N}$$
$$\frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

- We assume the initial conditions $(S(0), I(0), R(0))$ and the population size $N$ are known.
- The vector of model parameters is $\theta = (\beta, \gamma)$.
- The basic reproductive number is $\mathcal{R}_0 = \beta/\gamma$. Whenever $\beta/\gamma > 1$ then an outbreak occurs.

# Forward Solution versus Inverse Problem

- **Forward Problem**: Given some values of the true parameters then a numerical solution is computed. In symbols, given $\theta_0$ then $x(t, \theta_0)$ is computed.

# Forward Solution versus Inverse Problem

- **Forward Problem**: Given some values of the true parameters then a numerical solution is computed. In symbols, given $\theta_0$ then $x(t, \theta_0)$ is computed.
- **Inverse Problem**: Given $x(t, \theta_0)$ then $\theta_0$ is computed.

# Forward Solution versus Inverse Problem

- **Forward Problem**: Given some values of the true parameters then a numerical solution is computed. In symbols, given $\theta_0$ then $x(t, \theta_0)$ is computed.
- **Inverse Problem**: Given $x(t, \theta_0)$ then $\theta_0$ is computed.
- inverse problem $\approx$ parameter inference $\approx$ reverse engineering.

# Statistical Model for the Observation Process

- Observation = model output + measurement error

# Statistical Model for the Observation Process

- Observation = model output + measurement error
- Then, the statistical model for the observation process is

$$Y_i = I(t_i, \theta_0) + \mathcal{E}_i \quad \text{for } i = 1, \ldots, n.$$

# Statistical Model for the Observation Process

- Observation = model output + measurement error
- Then, the statistical model for the observation process is

$$Y_i = I(t_i, \theta_0) + \mathcal{E}_i \quad \text{for } i = 1, \ldots, n.$$

- The errors $\mathcal{E}_i$ are assumed to be random variables satisfying the following assumptions:
  (i) the errors $\mathcal{E}_i$ have mean zero, $E[\mathcal{E}_i] = 0$;

# Statistical Model for the Observation Process

- Observation = model output + measurement error
- Then, the statistical model for the observation process is

$$Y_i = I(t_i, \theta_0) + \mathcal{E}_i \quad \text{for } i = 1, \ldots, n.$$

- The errors $\mathcal{E}_i$ are assumed to be random variables satisfying the following assumptions:
  (i) the errors $\mathcal{E}_i$ have mean zero, $E[\mathcal{E}_i] = 0$;
  (ii) the errors $\mathcal{E}_i$ have finite common variance, $\text{var}(\mathcal{E}_i) = \sigma_0^2 < \infty$;

# Statistical Model for the Observation Process

- Observation = model output + measurement error
- Then, the statistical model for the observation process is

$$Y_i = I(t_i, \theta_0) + \mathcal{E}_i \quad \text{for } i = 1, \ldots, n.$$

- The errors $\mathcal{E}_i$ are assumed to be random variables satisfying the following assumptions:
  (i) the errors $\mathcal{E}_i$ have mean zero, $E[\mathcal{E}_i] = 0$;
  (ii) the errors $\mathcal{E}_i$ have finite common variance, $\mathrm{var}(\mathcal{E}_i) = \sigma_0^2 < \infty$;
  (iii) the errors $\mathcal{E}_i$ are i.i.d. (i.e., $\mathrm{cov}(\mathcal{E}_i, \mathcal{E}_j) = 0$ whenever $i \neq j$)

# OLS Example: Influenza in a Boarding School

- $$\frac{dS}{dt} = -\beta S \frac{I}{N}$$
$$\frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I,$$

  1. $N = 763$, $S(0) = 738$, $I(0) = 25$, $R(0) = 0$.
  2. The unknown model parameters are $\theta = (\beta, \gamma)$.

- Ordinary Least Squares (OLS):

$$\min_{\theta \in \Theta} \sum_{j=1}^{12} [y_j - I(t_j, \theta)]^2 .$$

# OLS Example: Influenza in a Boarding School

- The estimate $\hat{\theta}_{OLS}$ can be calculated applying one of several optimization algorithms (for example, Nelder-Mead simplex).

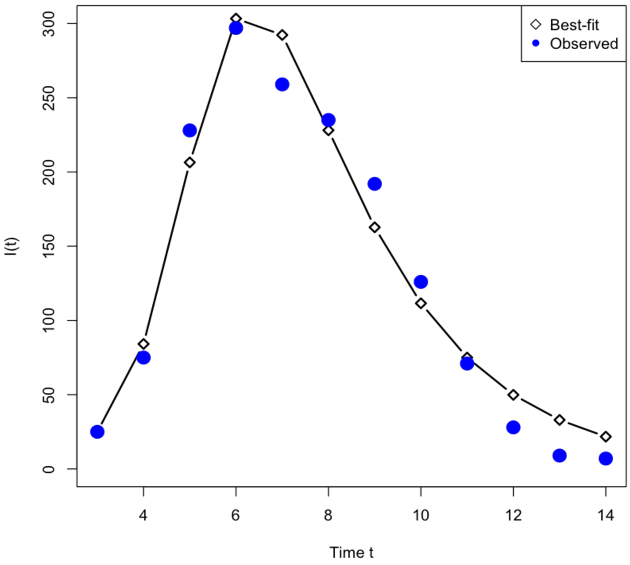
# OLS Example: Influenza in a Boarding School

- The estimate $\hat{\theta}_{OLS}$ can be calculated applying one of several optimization algorithms (for example, Nelder-Mead simplex).
- The statistical package R has a built-in function called `optim`. The GA library has a function called `ga`.

# OLS Example: Influenza in a Boarding School

- The estimate $\hat{\theta}_{OLS}$ can be calculated applying one of several optimization algorithms (for example, Nelder-Mead simplex).
- The statistical package R has a built-in function called `optim`. The GA library has a function called ga.
- $$\hat{\theta}_{OLS} = \arg\min_{\theta \in \Theta} \sum_{j=1}^{12} \left[ y_j - I(t_j; \theta) \right]^2 .$$

# OLS Example: Influenza in a Boarding School



# The Bootstrapping Method

1. First estimate $\hat{\theta}^0$ from the entire dataset, using OLS.
2. Using this estimate define the standard residuals:

$$\bar{r}_j = \sqrt{\frac{n}{n-p}}\left(y_j - I(t_j, \hat{\theta}^0)\right)$$

   for $j = 1, \ldots, n$. Then $\{\bar{r}_1, \ldots, \bar{r}_n\}$ are realizations of iid random variables $\bar{R}_j$ from the empirical distribution $\mathcal{F}_n$. Set $m = 0$.
3. Create a bootstrap sample of size $n$ using **simple random sampling with replacement** from $\{\bar{r}_1, \ldots, \bar{r}_n\}$ to form a bootstrap sample $\{\bar{r}_1^m, \ldots, \bar{r}_n^m\}$.
4. Create bootstrap sample points

$$y_j^m = I(t_j, \hat{\theta}^0) + \bar{r}_j^m,$$

   where $j = 1, \ldots, n$.
5. Obtain a new estimate $\hat{\theta}^{m+1}$ from the bootstrap sample $\{y_j^m\}$ using OLS.
6. Set $m = m + 1$ and repeat steps 3–5.

# The Bootstrapping Method

- Carry out the iterative process $M$ times where $M$ is large (e.g. $M = 1000$).

# The Bootstrapping Method

- Carry out the iterative process $M$ times where $M$ is large (e.g. $M = 1000$).
- R has the command `sample` for SRS.

# The Bootstrapping Method

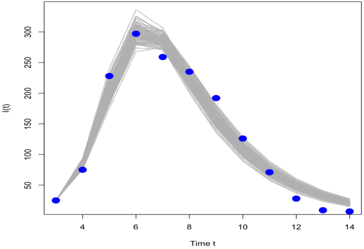
- Carry out the iterative process $M$ times where $M$ is large (e.g. $M = 1000$).
- R has the command `sample` for SRS.
- Calculate mean, covariance matrix, and standard errors using the formulae:

$$\hat{\theta}_{boot} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}^m$$

$$\mathrm{Cov}(\hat{\theta}_{boot}) = \frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{\theta}^m - \hat{\theta}_{boot}\right)^T \left(\hat{\theta}^m - \hat{\theta}_{boot}\right)$$

$$SE_k(\hat{\theta}_{boot}) = \sqrt{\mathrm{Cov}(\hat{\theta}_{boot})}$$

# The Bootstrapping Method




| Parameter | Estimate | SE |
|---|---|---|
| $\beta$ | 1.8191 | 0.0608 |
| $\gamma$ | 0.4703 | 0.0173 |
| $\mathcal{R}_0 = \beta/\gamma$ | 3.8680 | N/A |