Fitting time series models, Part 2

CIHR Course Week 3

Isabel (Izzie) Fulcher

isabel_fulcher@hms.harvard.edu

Teaching Objectives

- Quick recap of time series modeling from Part 1
- Seasonality and time trends
- Syndromic surveillance
  - Indicators, baseline, and evaluation periods
  - Prediction intervals

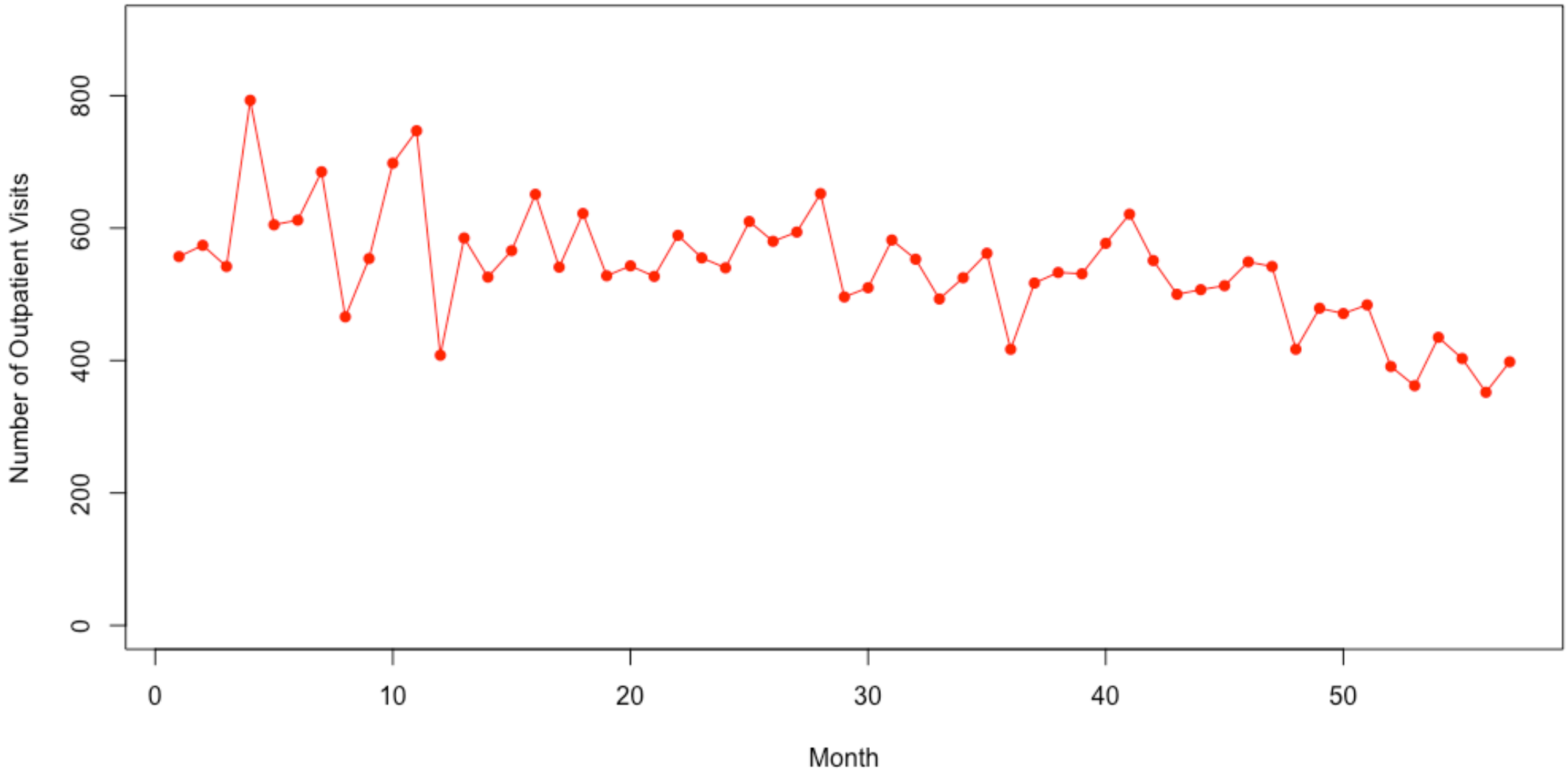
Part 1 Recap

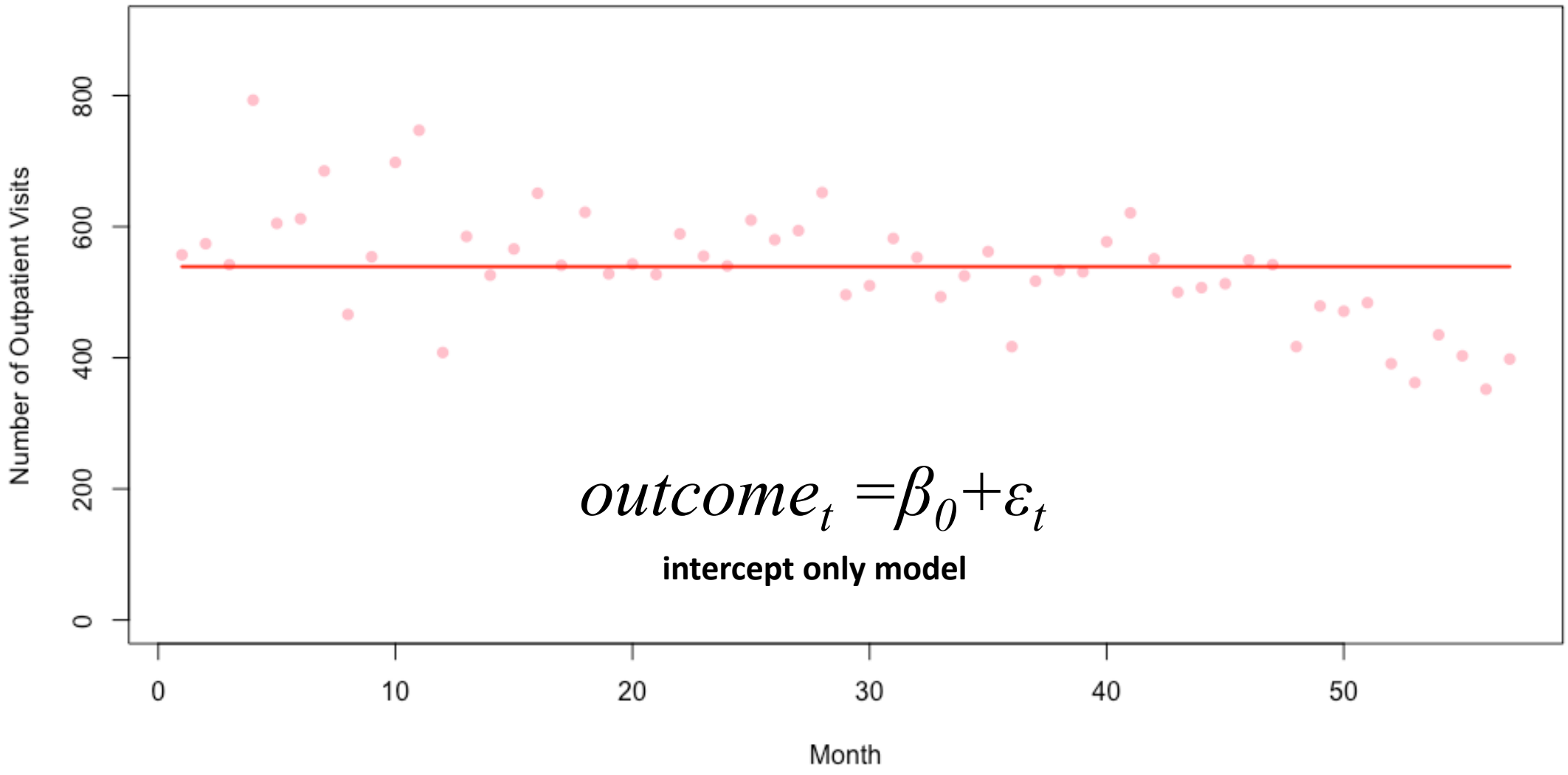
Research questions with time series data

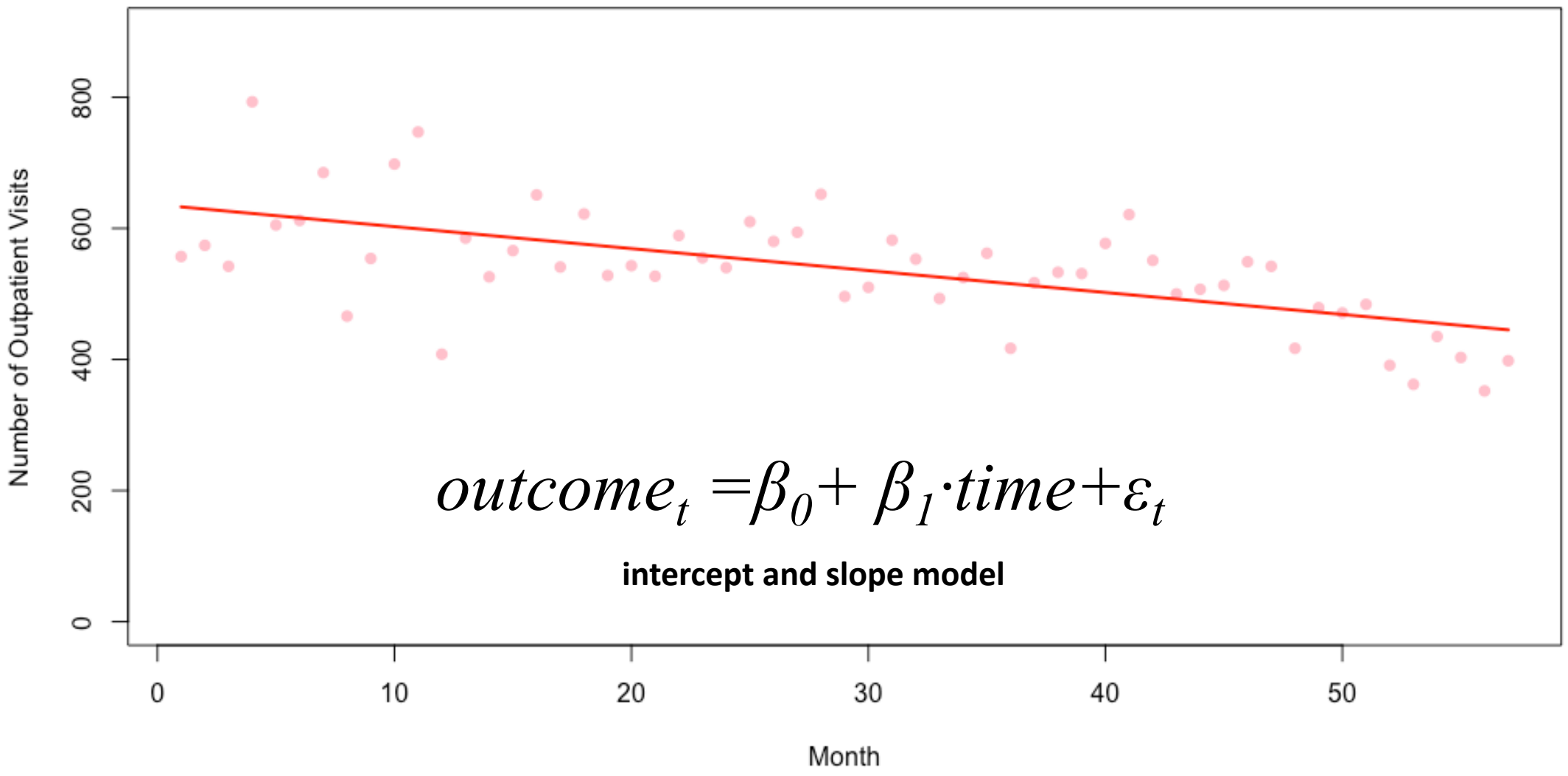
- **Describe** the behavior of an indicator over time
  - Is infant mortality decreasing over time? By how much?
  - What are the seasonal differences in malaria cases?

- **Detect** deviations from expected in an indicator
  - Is a region experiencing higher than expected cases of diarrhea?
  - Is the number of health facility deliveries lower than expected?

- **Measure** the impact of an intervention on an indicator
  - After social distancing measures were put in place, how many fewer COVID-19 cases were there?

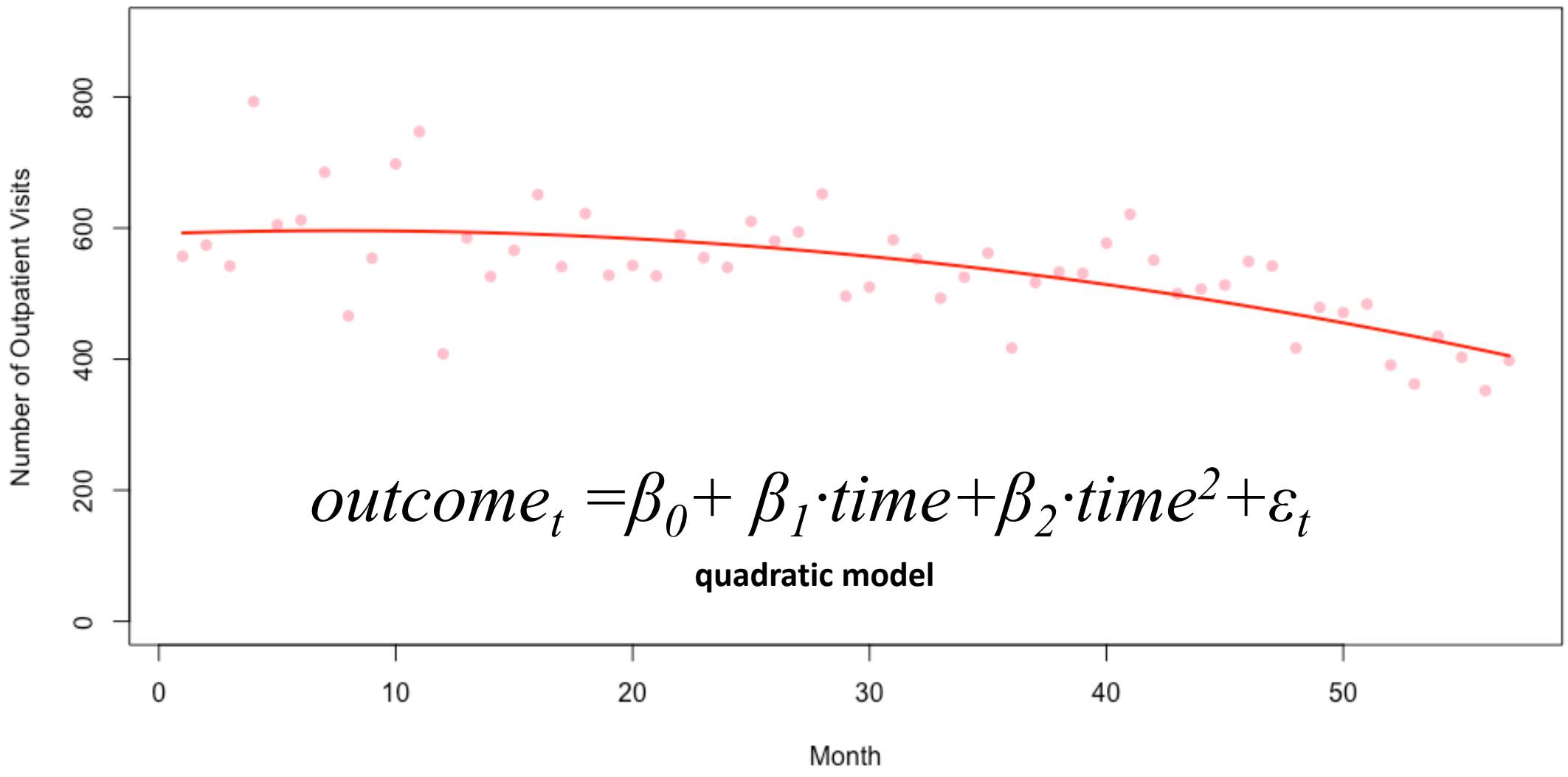

Number of Outpatient Visits
0
200
400
600
800
0
10
20
30
40
50
Month


Number of Outpatient Visits
0
200
400
600
800
0
10
20
30
40
50
Month
$outcome_t = \beta_0 + \varepsilon_t$
intercept only model



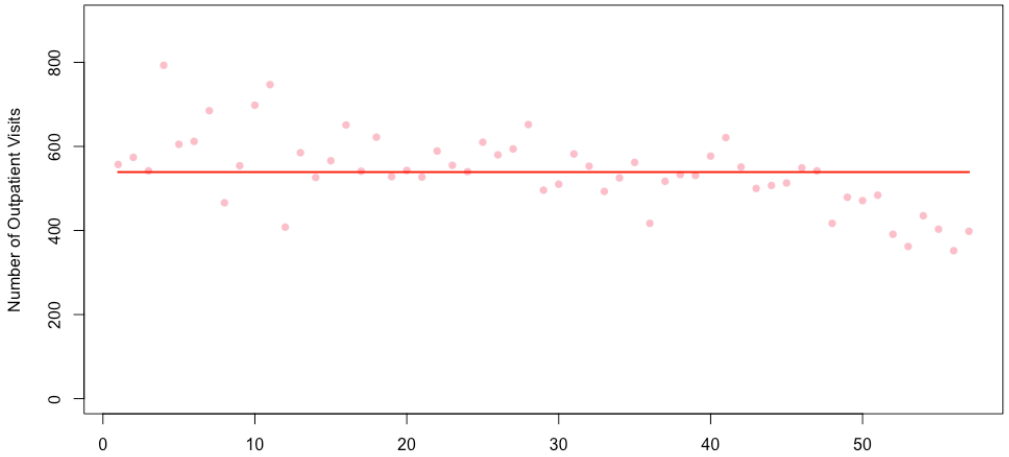
$$outcome_t = \beta_0 + \beta_1 \cdot time + \varepsilon_t$$

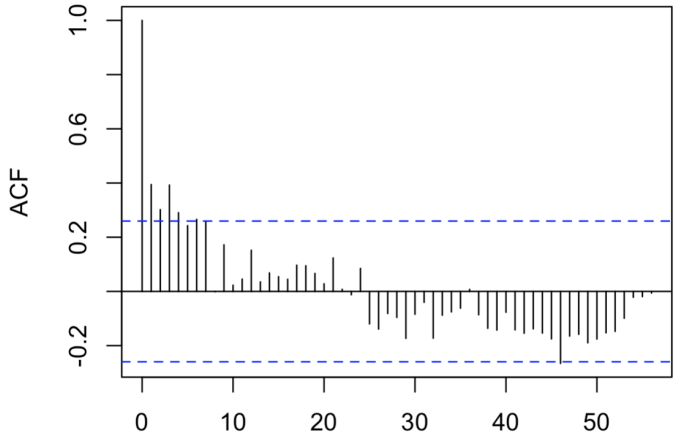
intercept and slope model


Number of Outpatient Visits
0
200
400
600
800
0
10
20
30
40
50
Month
$outcome_t = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2 + \varepsilon_t$
quadratic model

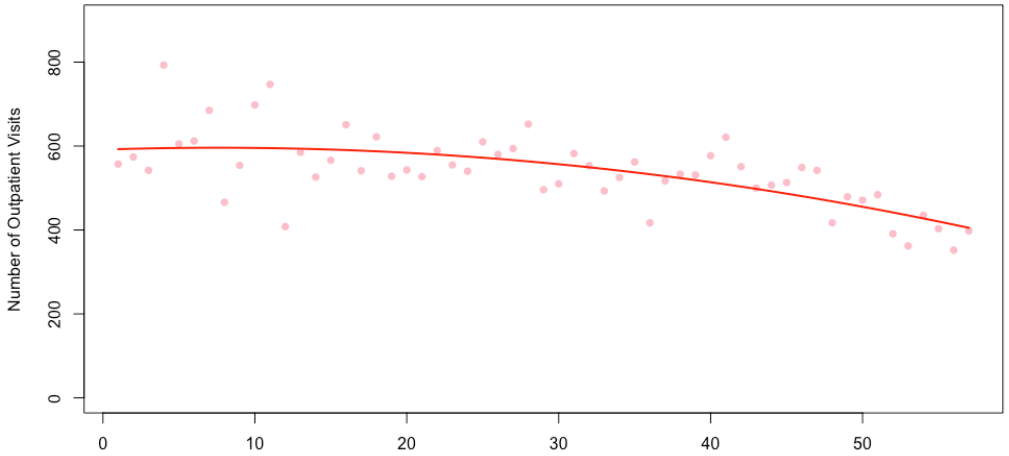
Model diagnostics: autocorrelation

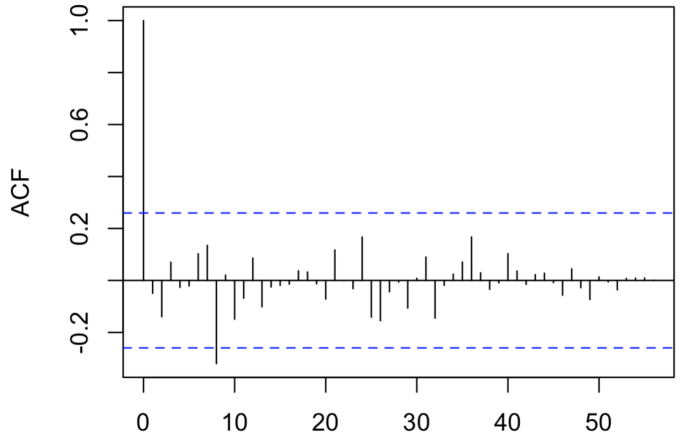
intercept only model





quadratic model





Seasonality and time trends

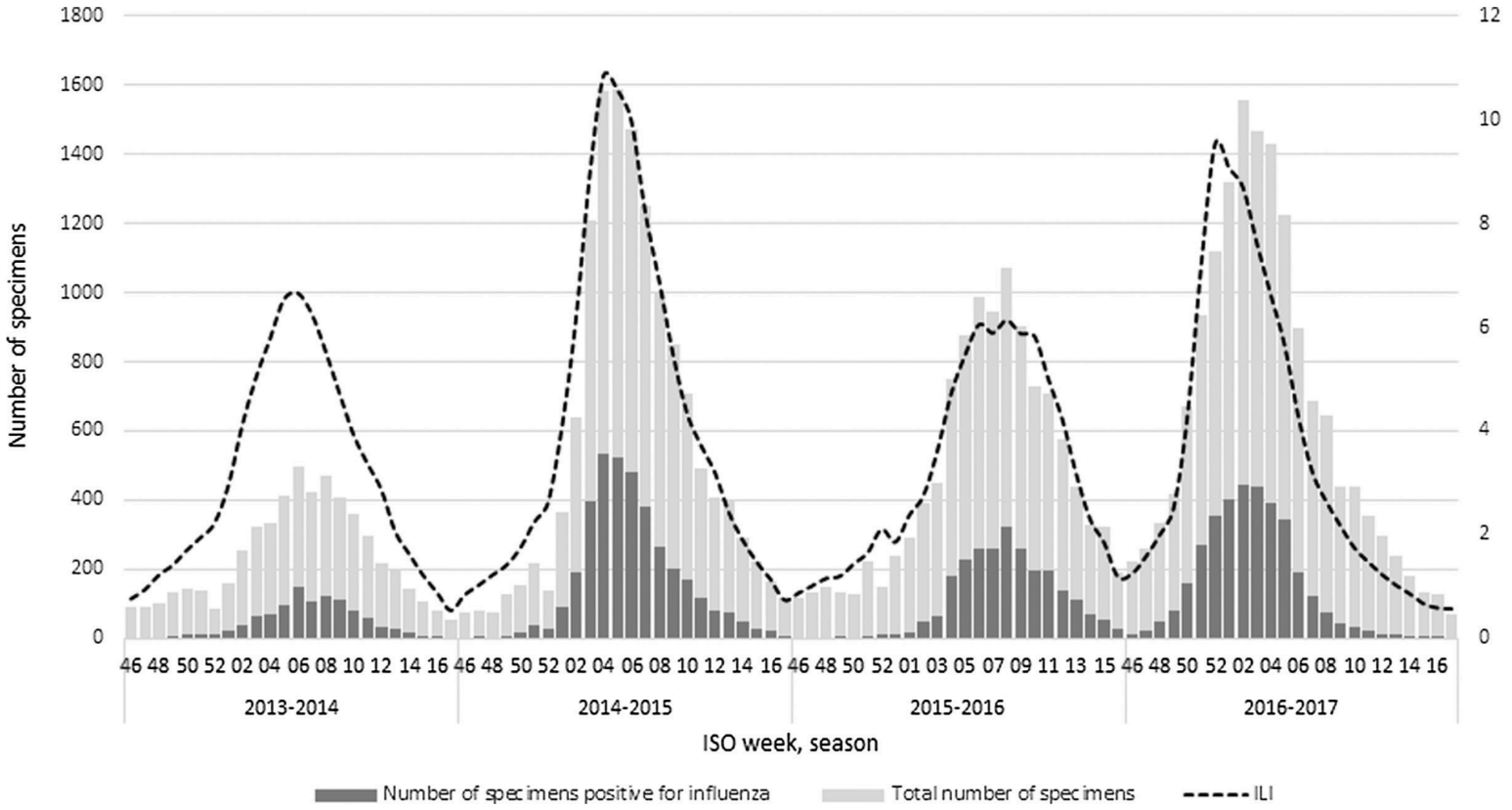
Time series patterns

- **Seasonal:** Outcomes are impacted by seasonal factors (season, month, day of the week)
  - Fixed with known frequency
  - Easily mistaken with fluctuations (must be related to some aspect of calendar time)

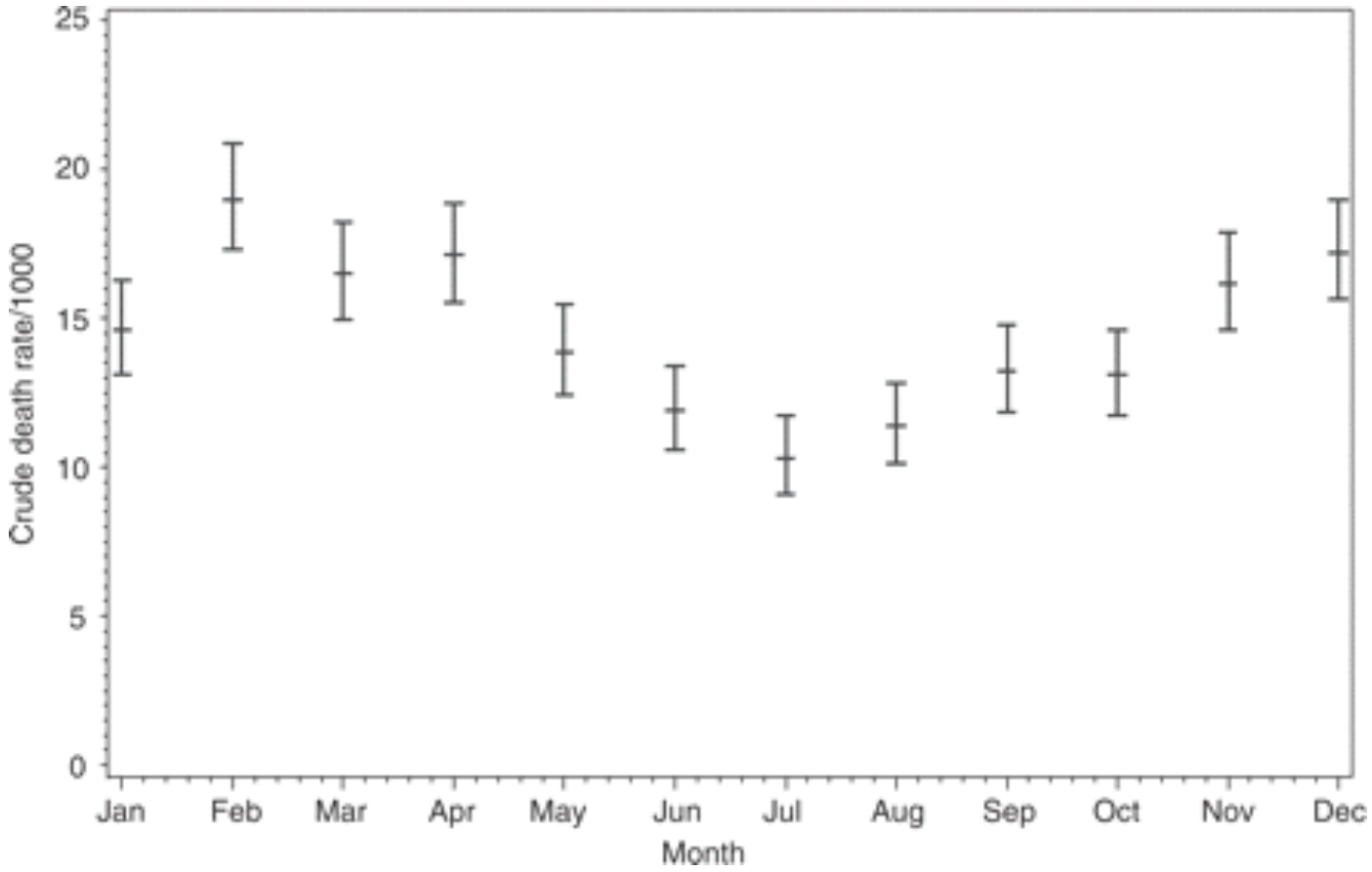
- **Trend:** Long-term increase or decrease in the data.
  - Trends can change direction over time
  - Trends do not need to be linear

Source: Forecasting: Principles and Practice (https://otexts.com/fpp2/)

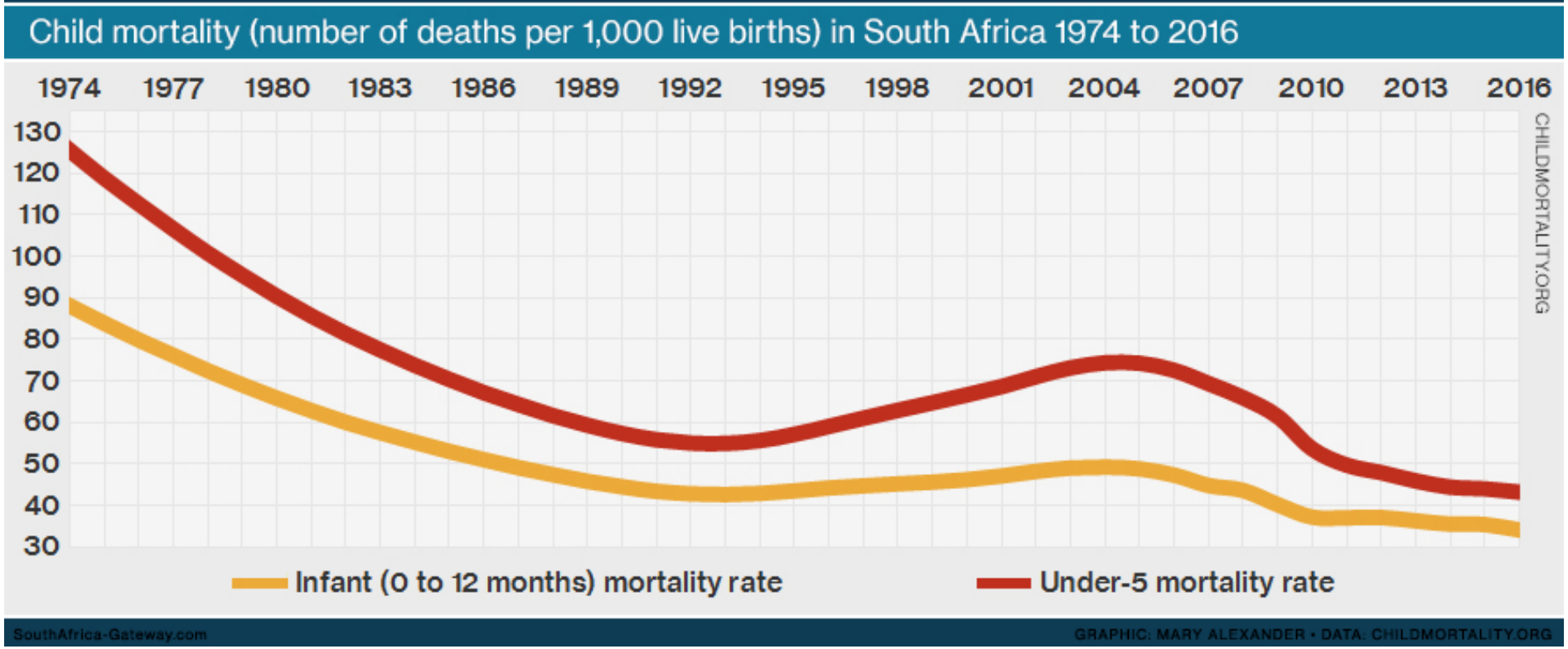
Influenza in Italy
(Rosano et al., 2019)



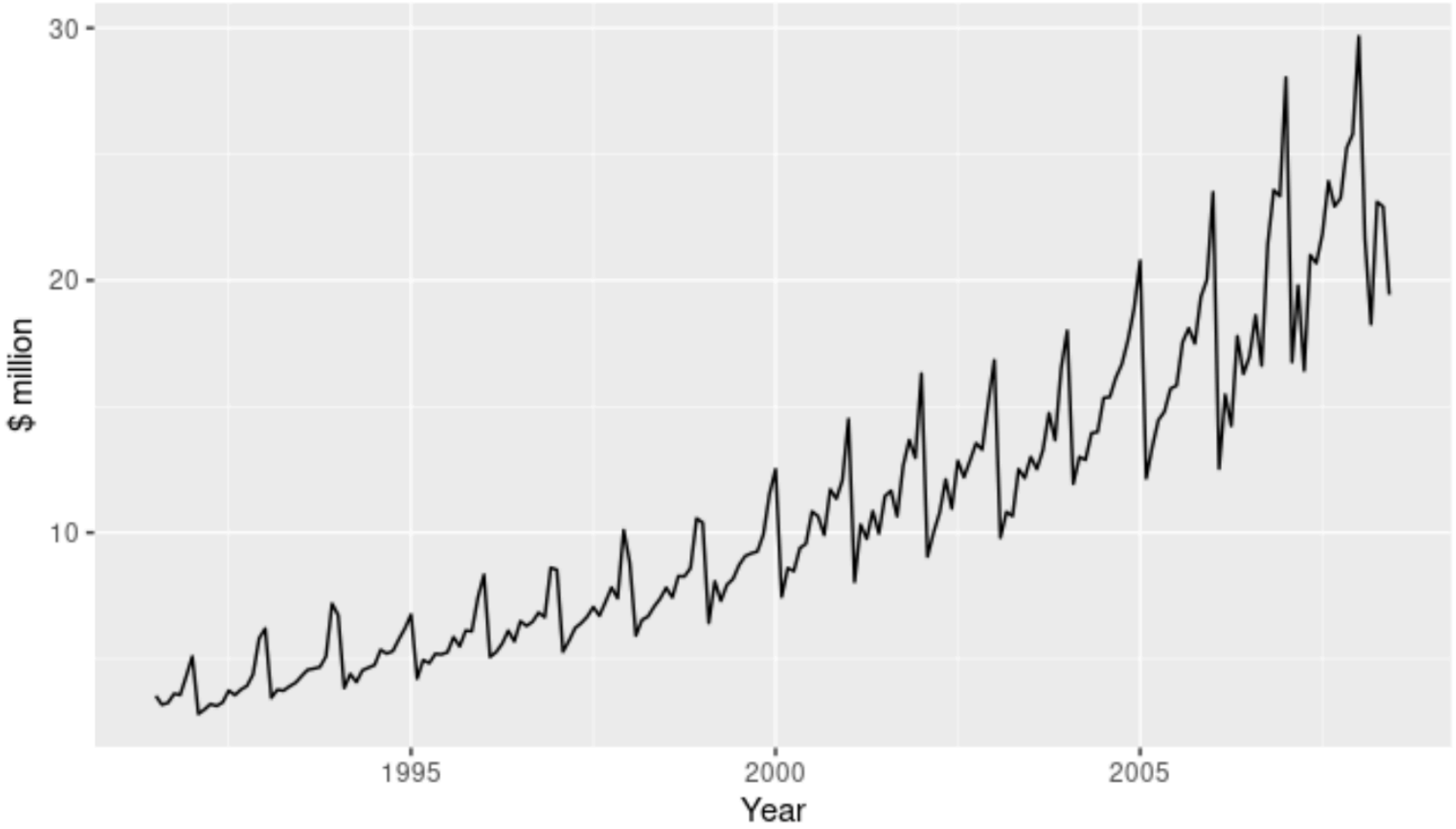
Death rates in Burkina Faso from 1993-2001 (Kynast-Wolf et al., 2005)



Child mortality from 1974 to 2016



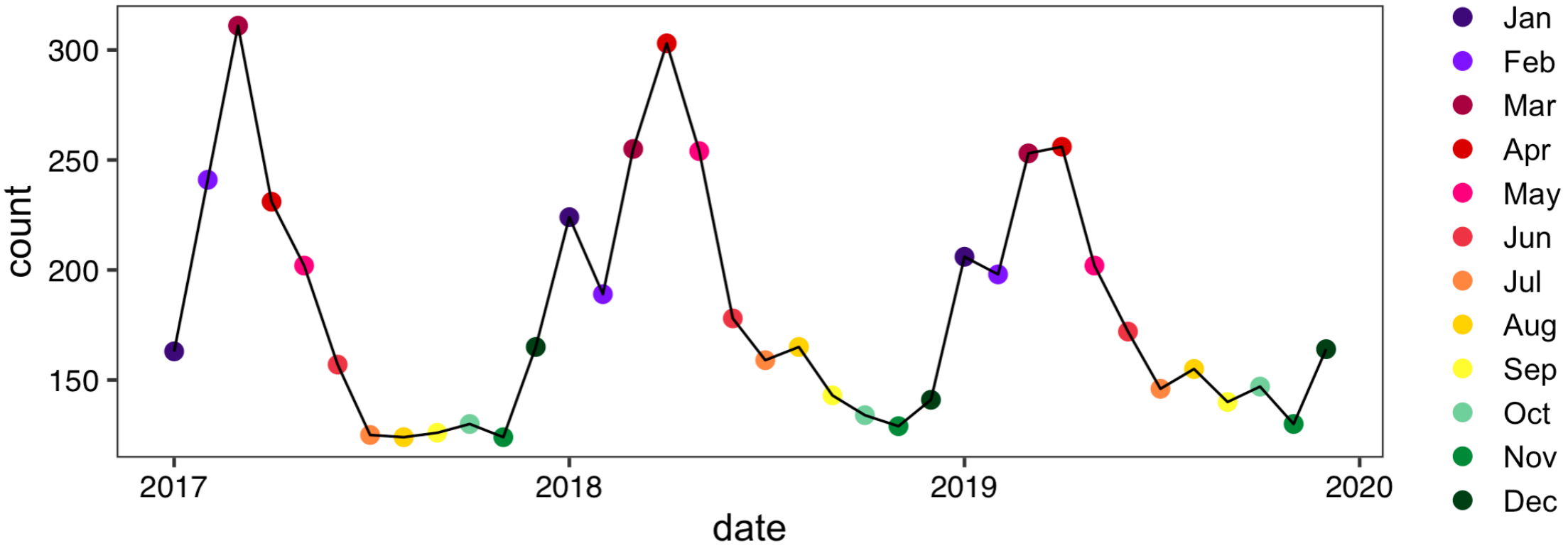
Antidiabetic drug sales in Australia



Source: Forecasting: Principles and Practice (https://otexts.com/fpp2/)

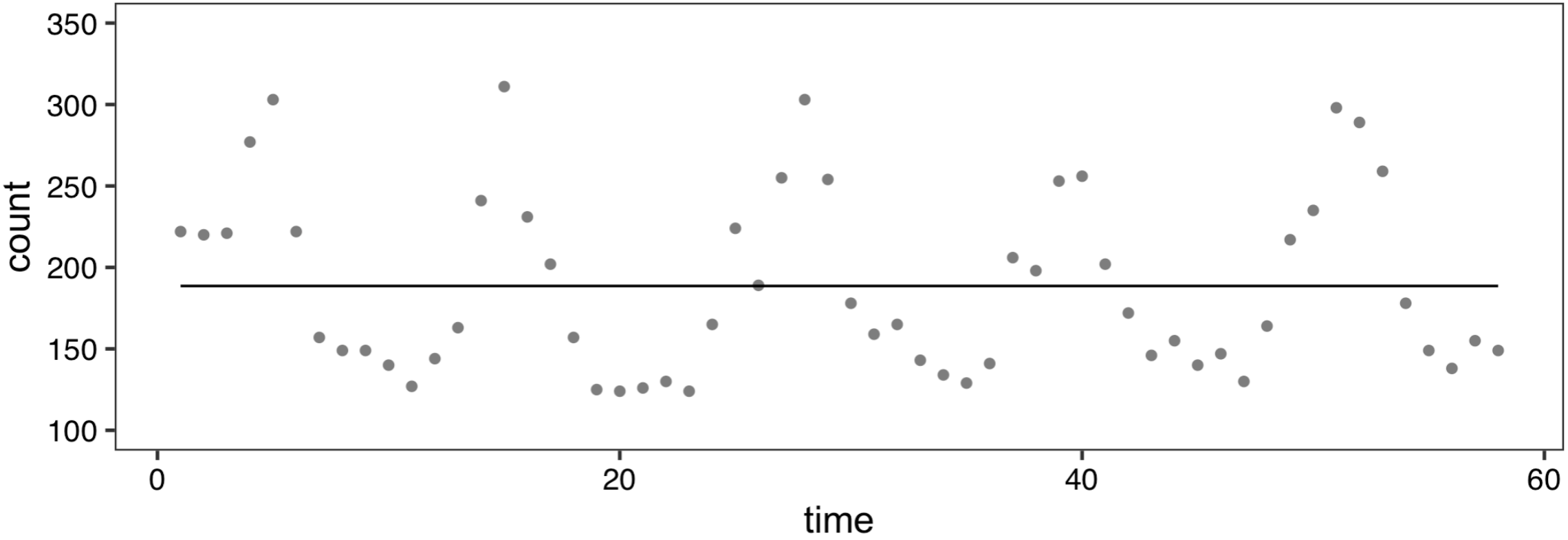
Activity: Malaria cases in Malawi facility

Please go to www.menti.com with 4897 0509



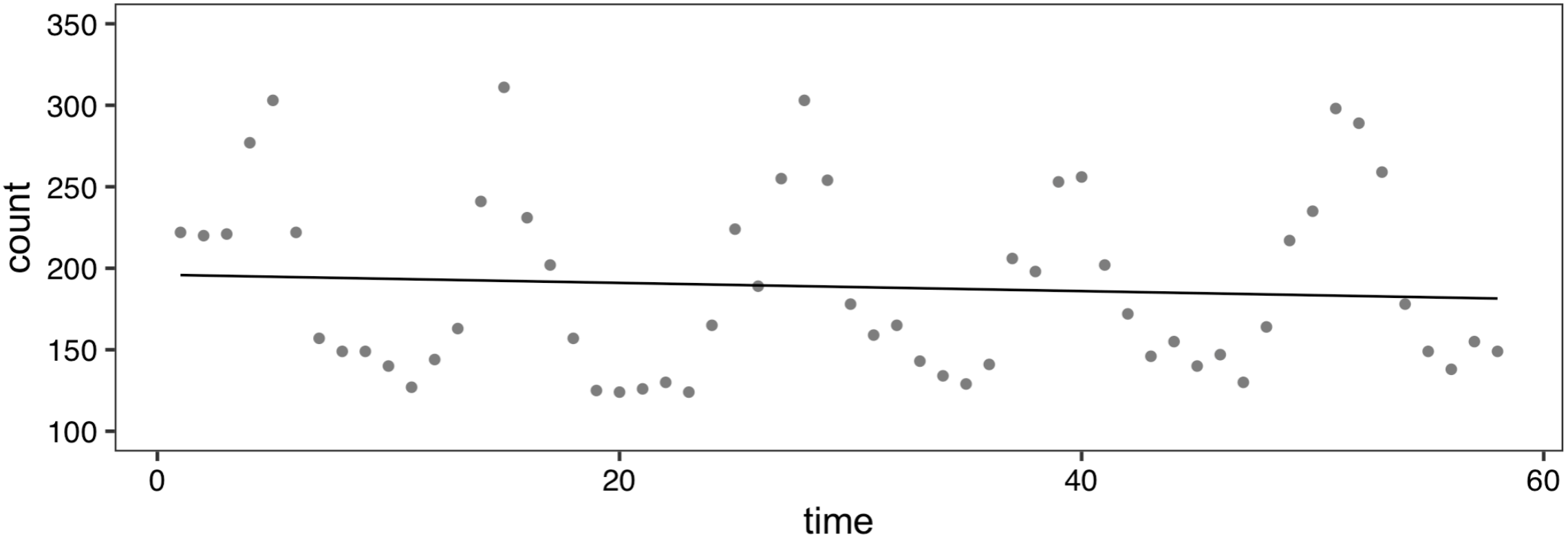
Malaria cases in Malawi facility

$$outcome_t = \beta_0 + \varepsilon_t$$



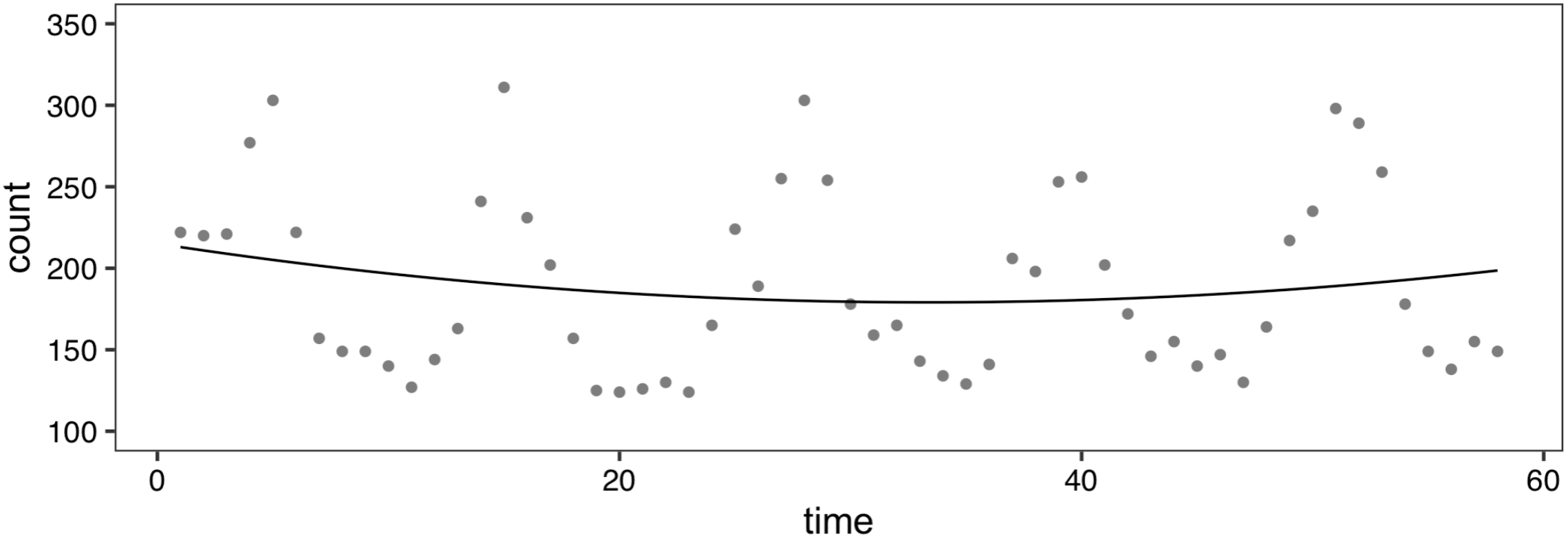
Malaria cases in Malawi facility

$$outcome_t = \beta_0 + \beta_1 \cdot time + \varepsilon_t$$



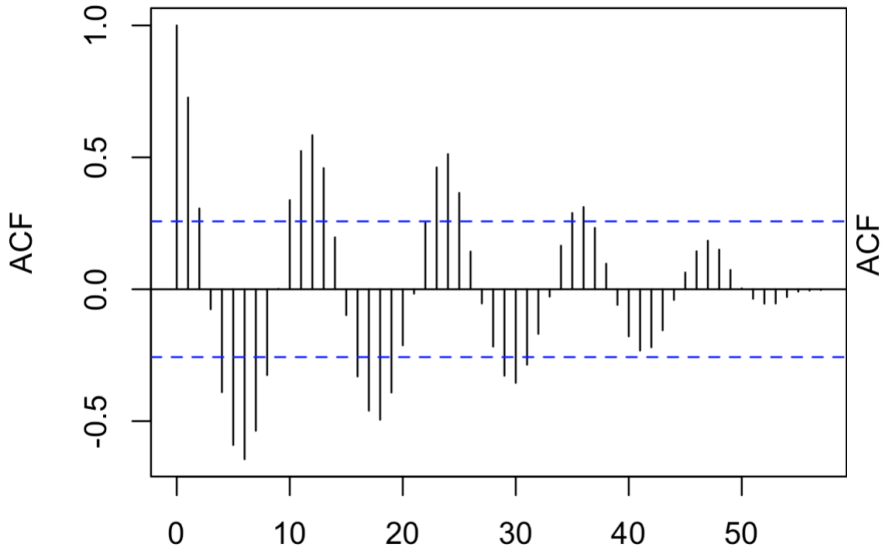
Malaria cases in Malawi facility

$$outcome_t = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2 + \varepsilon_t$$

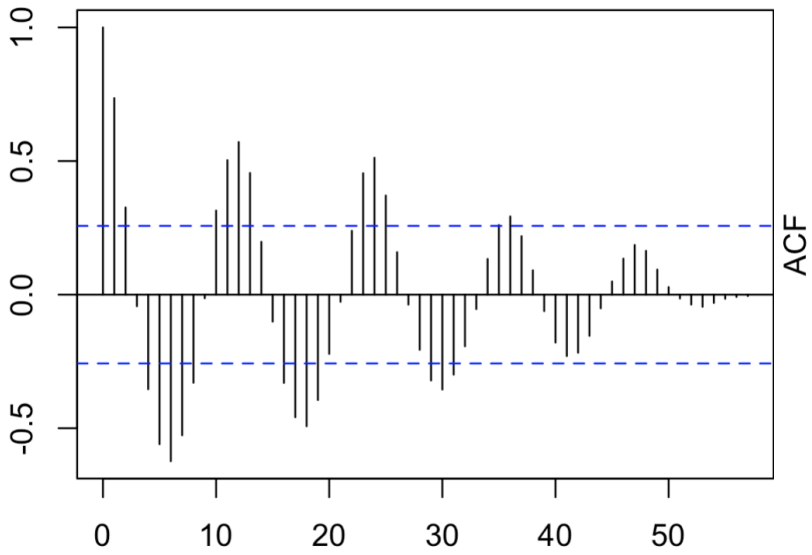


Model diagnostics: autocorrelation

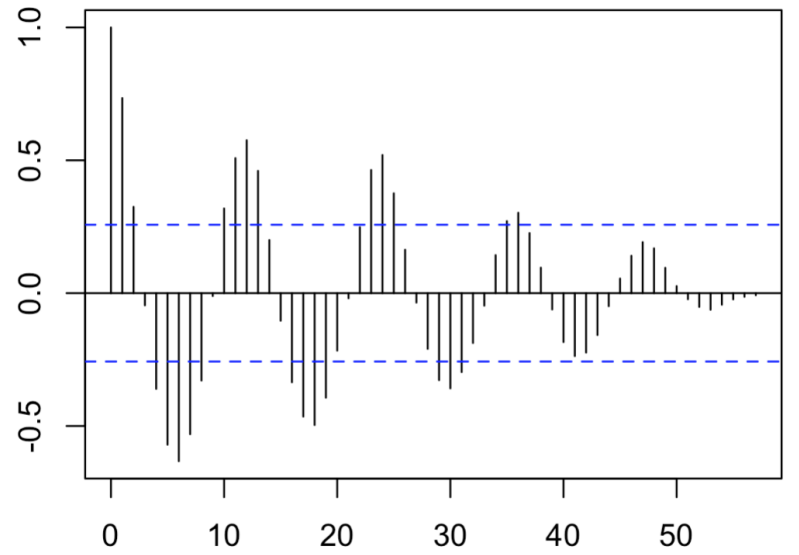
intercept only model



intercept and slope model



quadratic model



None of these models see any improvement over the other!
Need another method to capture "seasonality"!

How to deal with seasonality and trend?

- **Seasonality:** Fourier series (or harmonic functions) are popular ways to deal with seasonality in time series models

$$x_{1,t} = \sin\left(\frac{2\pi t}{m}\right), x_{2,t} = \cos\left(\frac{2\pi t}{m}\right), x_{3,t} = \sin\left(\frac{4\pi t}{m}\right),$$

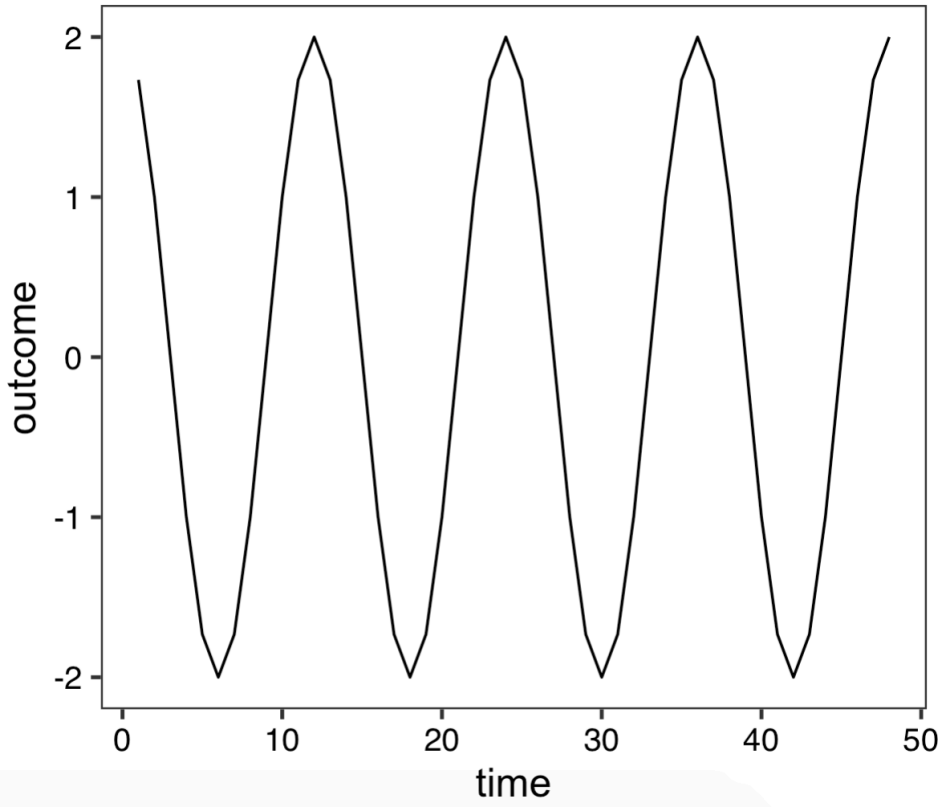
$$x_{4,t} = \cos\left(\frac{4\pi t}{m}\right), x_{5,t} = \sin\left(\frac{6\pi t}{m}\right), x_{6,t} = \cos\left(\frac{6\pi t}{m}\right),$$

- **Trends:** Add a linear term(s) to capture yearly trend
  - Will only capture linear decreases or increases over time
  - If changes are not linear, dummy variables or more complex transformations (log, quadratic, etc.) can be used

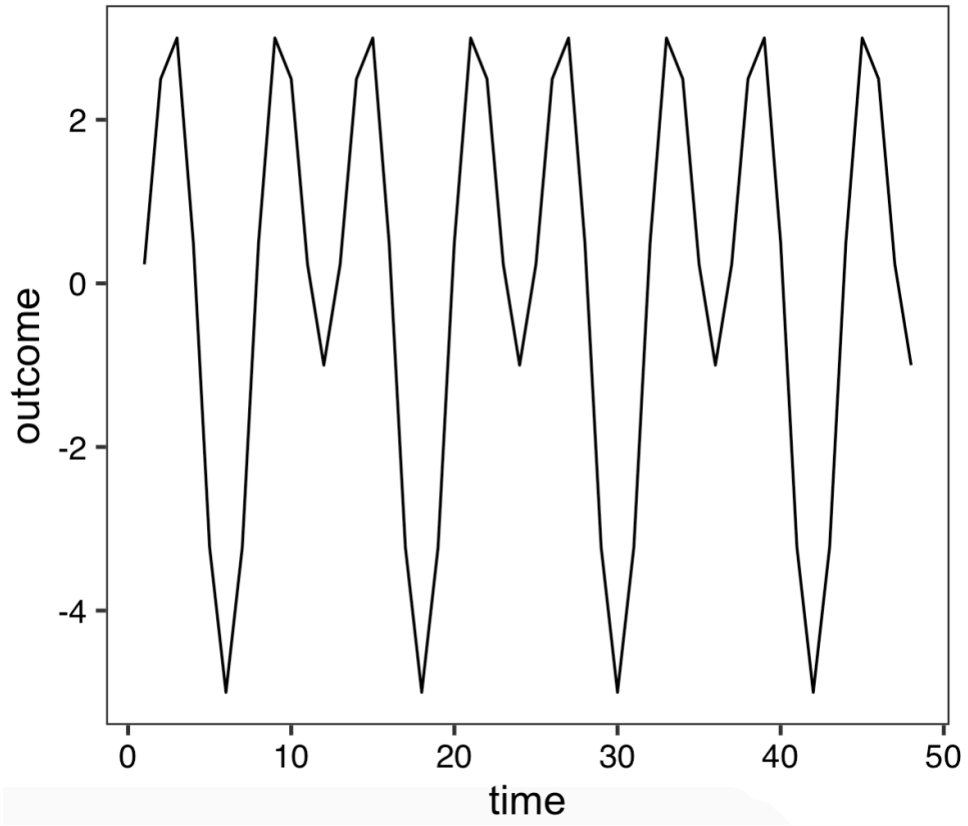
Source: Forecasting: Principles and Practice (https://otexts.com/fpp2/)

Fourier series (*examples*)

$sin(2\pi t/12) + cos(2\pi t/12)$

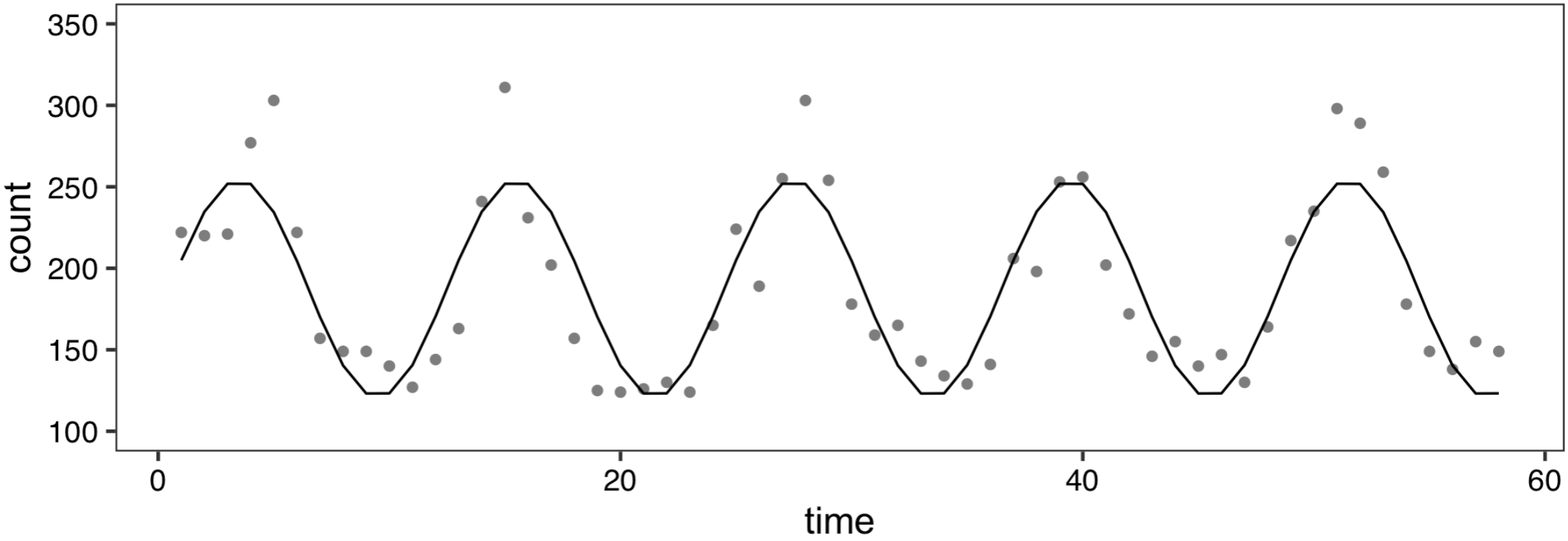


$sin(2\pi t/12) + cos(2\pi t/12) + sin(4\pi t/12) + cos(4\pi t/12)$

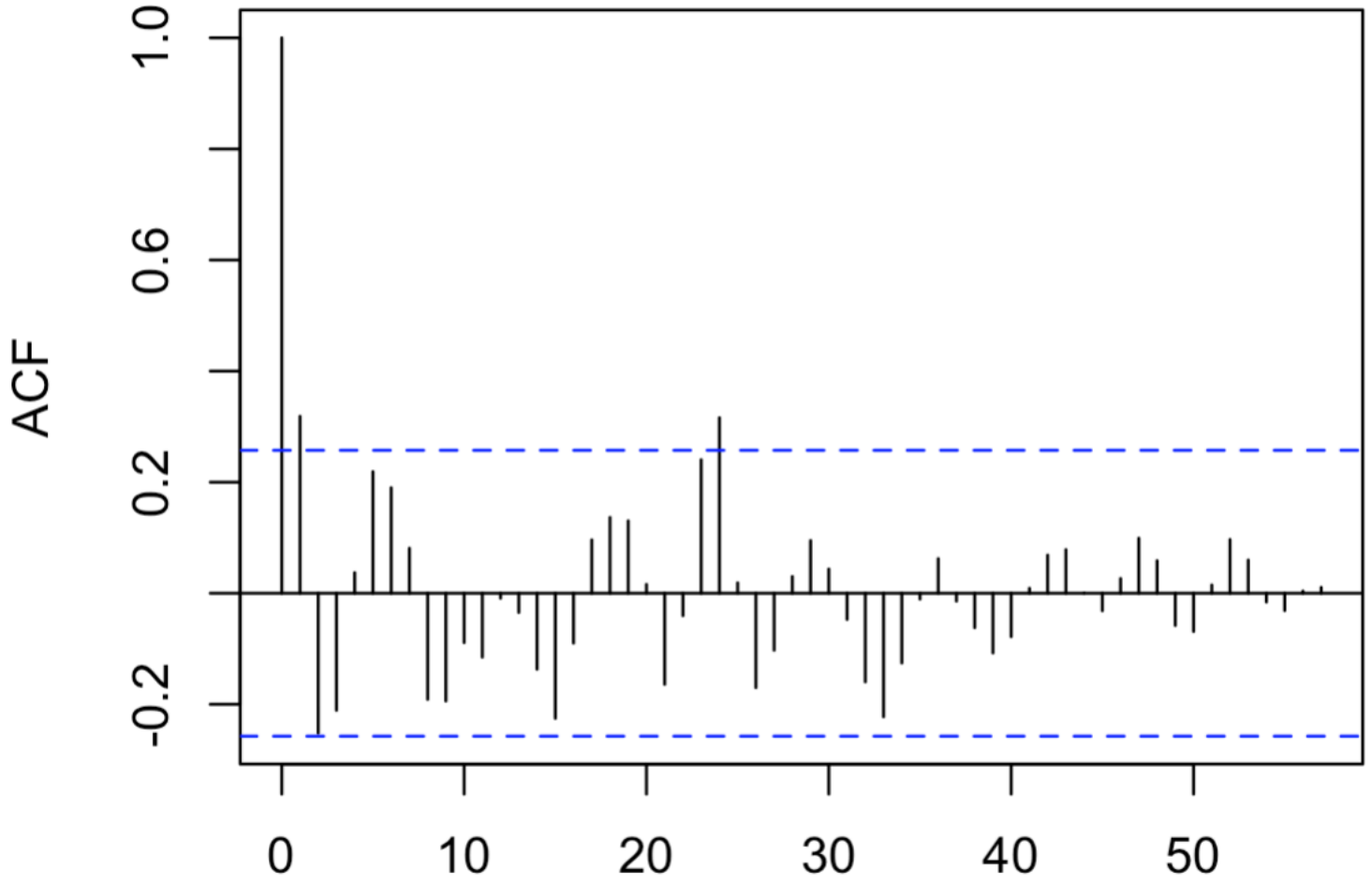


Malaria cases in Malawi facility

$$outcome_t = \beta_0 + \beta_1 \cdot sin(2\pi t/12) + \beta_2 \cdot cos(2\pi t/12) + \varepsilon_t$$



Model diagnostics: autocorrelation



Syndromic surveillance

Research questions with time series data

- **Describe** the behavior of an indicator over time
  - Is infant mortality decreasing over time? By how much?
  - What are the seasonal differences in malaria cases?

- **Detect** deviations from expected in an indicator
  - Is a region experiencing higher than expected cases of diarrhea?
  - Is the number of health facility deliveries lower than expected?

- **Measure** the impact of an intervention on an indicator
  - After social distancing measures were put in place, how many fewer COVID-19 cases were there?

Detecting deviations from expected

Syndromic surveillance: is the process of monitoring data on symptoms or outcomes related to a disease as a way to **detect** areas that might be affected by the disease.

Pros:

- Provide "warnings" for local areas.
- Uses existing data.
- Process can be automated.

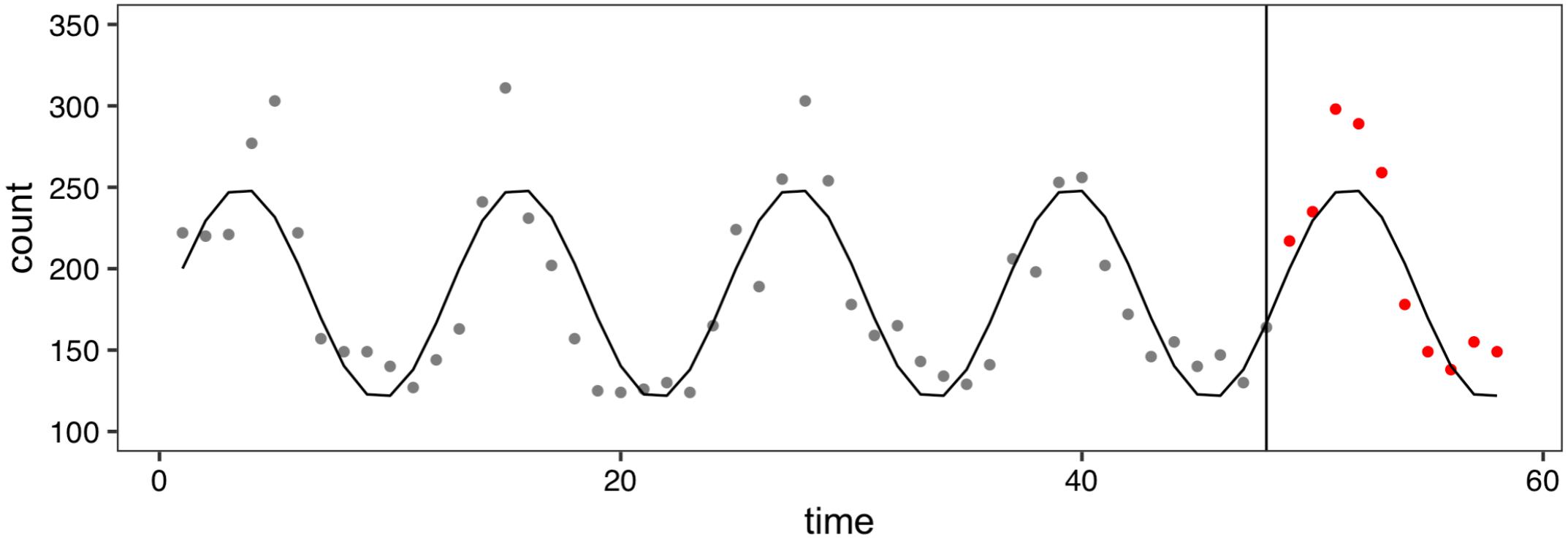
Cons:

- Not as good as monitoring the disease directly.
- Changes could be explained by **other factors** that should be considered.

How to conduct syndromic surveillance

- **Step 1.** Choose relevant indicators to follow over time (*Session 1*) and format data for analysis (*Session 4*)
- **Step 2.** Determine baseline and evaluation periods
- **Step 3.** Fit time series model to baseline period (*Sessions 2 & 3*)
- **Step 4.** Using the model from Step 3, calculate deviations from expected in the evaluation period (*Session 3*)
- **Step 5.** Produce interpretable visualizations (*Session 5*)

Group activity: *how should we calculate deviations from expected?*



Group activity: *how should we calculate deviations from expected?*

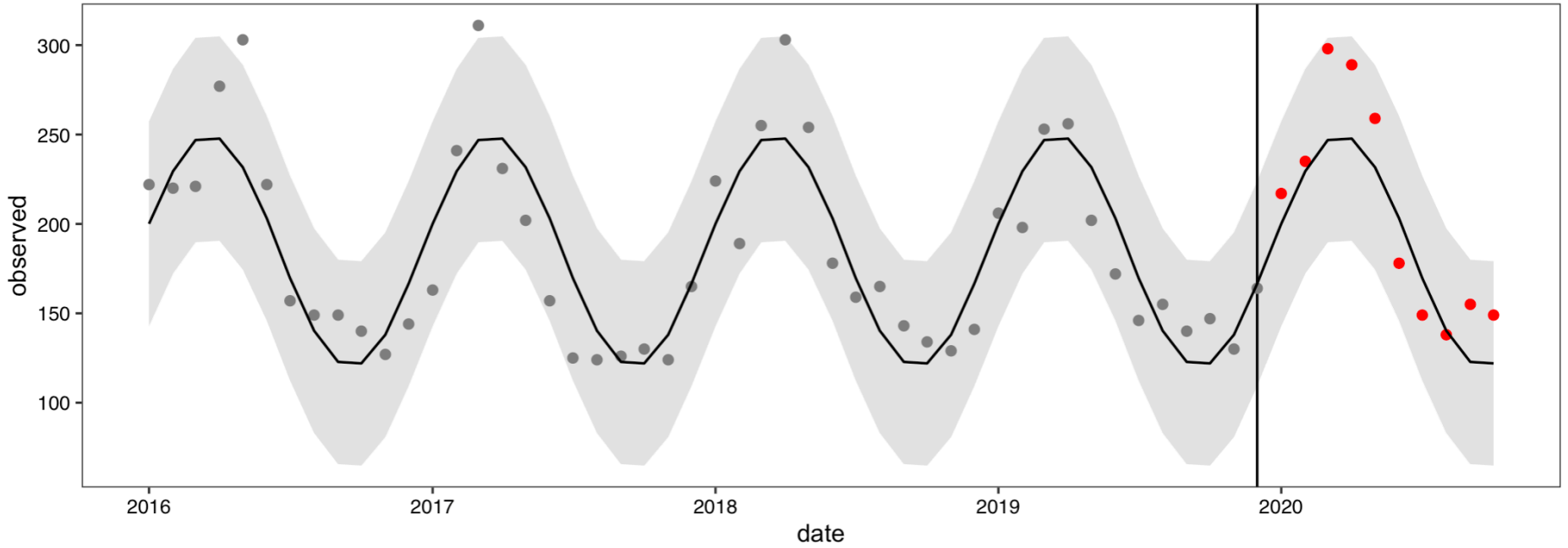
1. Are the number of malaria cases higher or lower than expected in January 2020? By how many malaria cases?

2. In the evaluation period, are the observed numbers of malaria cases systematically higher than expected, lower than expected, or is there no discernable pattern?

3. Does using *predicted cases - observed cases* to identify deviations make sense? How could this measure be improved?

Your group should answer these questions on a Google Form in a Breakout Room.

Prediction intervals



Prediction interval: the range of values where a future individual observation (monthly count) is likely to fall.

Prediction intervals are different than confidence intervals, because they correspond to an individual value, not the mean.

Lab activity: Syndromic Surveillance of Acute Respiratory Infection Cases at Liberia facility

- Lab will be taught by Don today

Next lecture: Cleaning the Data

- Extracting data
- Setting up data in usable format
- Identifying outliers
- Dealing with missing data
- Automation