

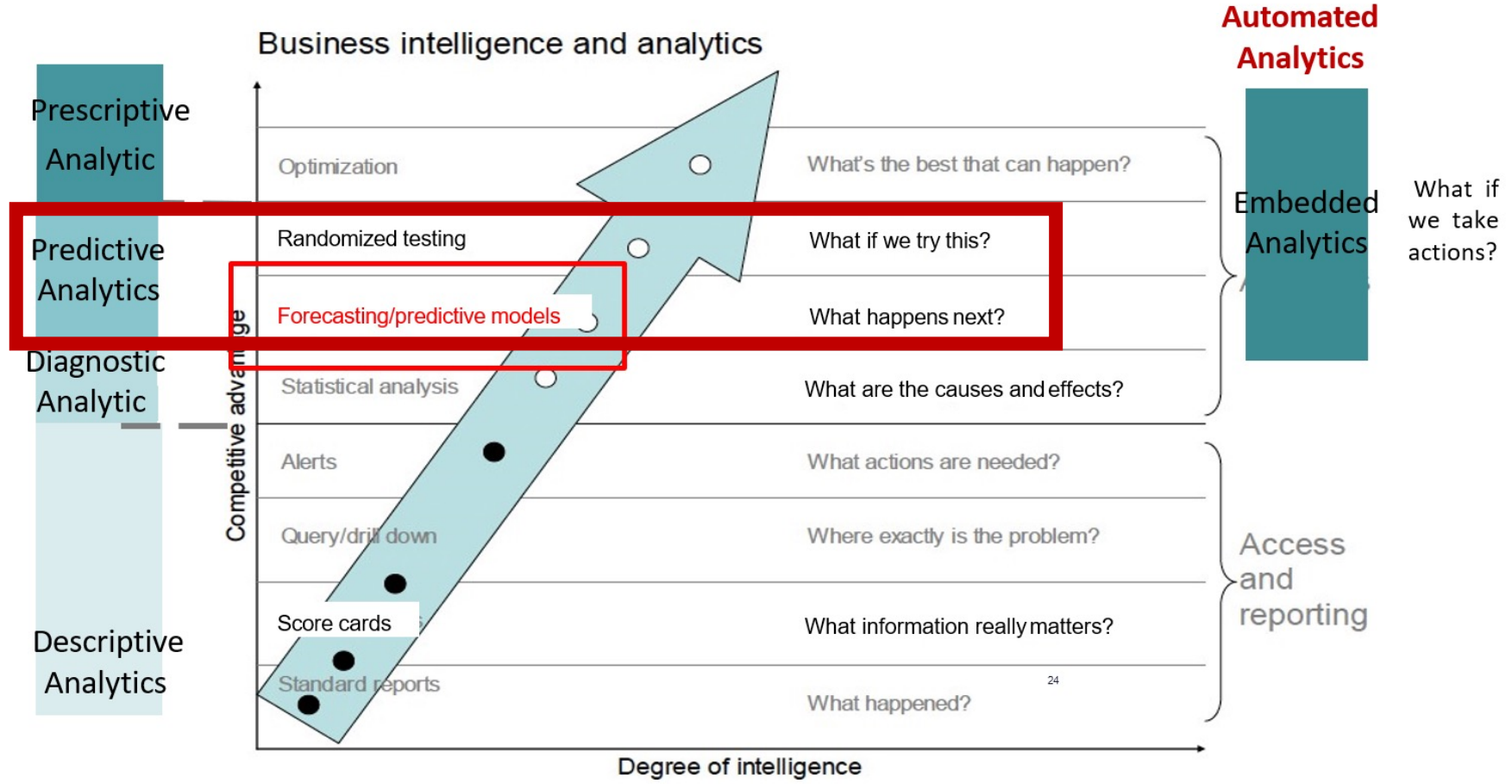


การวิเคราะห์ข้อมูลเชิงพยากรณ์ (Predictive Analytics)

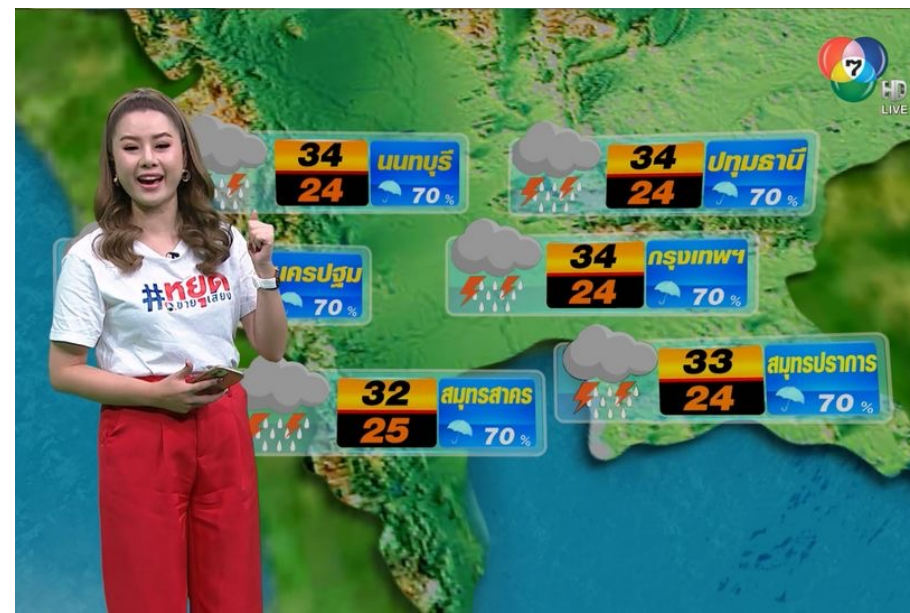
ศุภณัฐ วงศานุพัทธ์

กลุ่มพัฒนาระบบข้าราชการและเฝ้าระวังโรคไม่ติดต่อ (iNCD), กองระบาดวิทยา, กรมควบคุมโรค

ระดับชั้นของการวิเคราะห์ข้อมูล

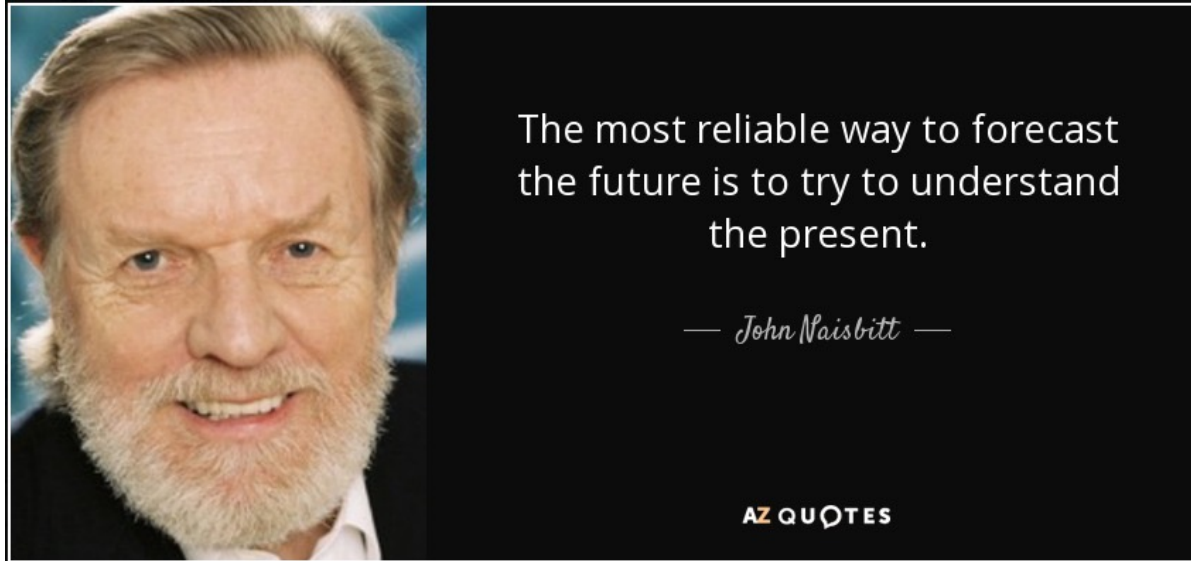


What is Forecasting?



What is Forecasting?

- **Forecasting** is the process of making **predictions of the future based on past and present data** and most commonly by analysis of trends.
- การใช้อดีตและปัจจุบันทำนายอนาคต





กฎของการพยากรณ์

- **Law 1** – การพยากรณ์ *ส่วนมากจะผิด (แต่ดีกว่าก็ิน!)*
 - Perfect prediction is impossible
 - Simple quantitative forecasting technique will generally *outperform* unstructured intuitive assessment.
- **Law 2** – การพยากรณ์ที่ *แยกย่อย* จะ *แม่นยำกว่าแบบหยาบ*
 - Due to variability pooling, aggregate forecast can cancel out unwanted variations
 - Aggregate the past demand values of the same product across multiple locations, or aggregate the past demand values of the same product in terms of time buckets
- **Law 3** – ยิ่งพยากรณ์ไป *ไกล* ยิ่ง *แม่นยำน้อยลง*

จะพยากรณ์โรคเมื่อใด

Disease Importance (ความสำคัญของโรค)

- Frequency
- Severity
- Impact (QOL, Social, economic)
- Public/political concern

Modifiability (ความเป็นไปได้ในการ ปรับเปลี่ยน)

- Preventable disease
- Available effective intervention
- Acceptable cost

Predictability (ความเป็นไปได้ในการ พยากรณ์)

- Available data
- Available technique
- Available technology
- Available experts and experiences (both content and technique)



องค์ประกอบที่ดีของการพยากรณ์

- Compatible (สอดคล้อง)
- Meaningful (มีความหมาย)
- Useful time horizon (ช่วงเวลาที่ยากรณ์มีประโยชน์)
- Reliable (เชื่อถือได้)
- Accurate (แม่นยำ)
- Easy to understand and use (ง่ายต่อการเข้าใจและใช้งาน)

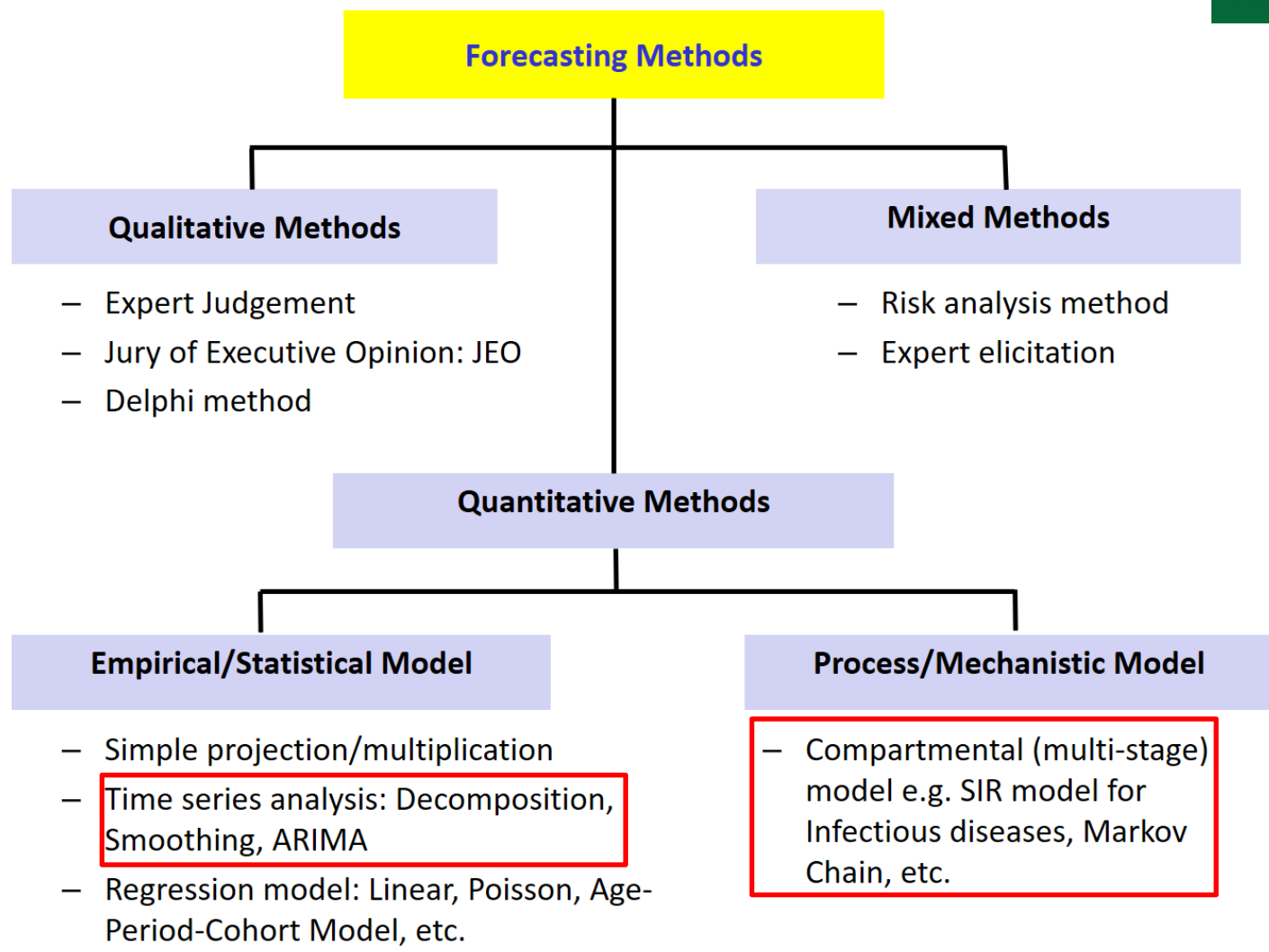


ประโยชน์ของการพยากรณ์โรค

- เตรียมความพร้อม และ วางแผนทางทรัพยากร
- ช่วยการตัดสินใจเชิงนโยบาย
- ประเมินประสิทธิภาพของนโยบาย
- แจ้งเตือนการระบาด
- ควบคุมการระบาด
- สร้างความตระหนัก



วิธีการพยากรณ์โรค





การพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting)

What are time series data?

- **Cross sectional data**

- **Unidimensional on time** (but possibly covering groups of populations) (ณ ช่วงเวลาใด ช่วงเวลาหนึ่ง)

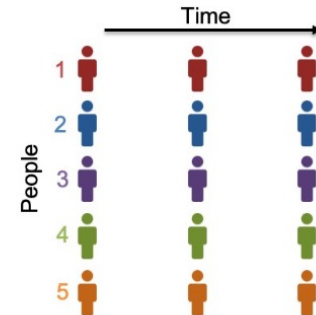
Single measurement data



- **Panel data**

- Longitudinal multidimensional data involving measurements **over time on the same individuals** (ติดตามคนๆเดิมไปในหลายๆช่วงเวลา)

Panel Data



What are time series data?

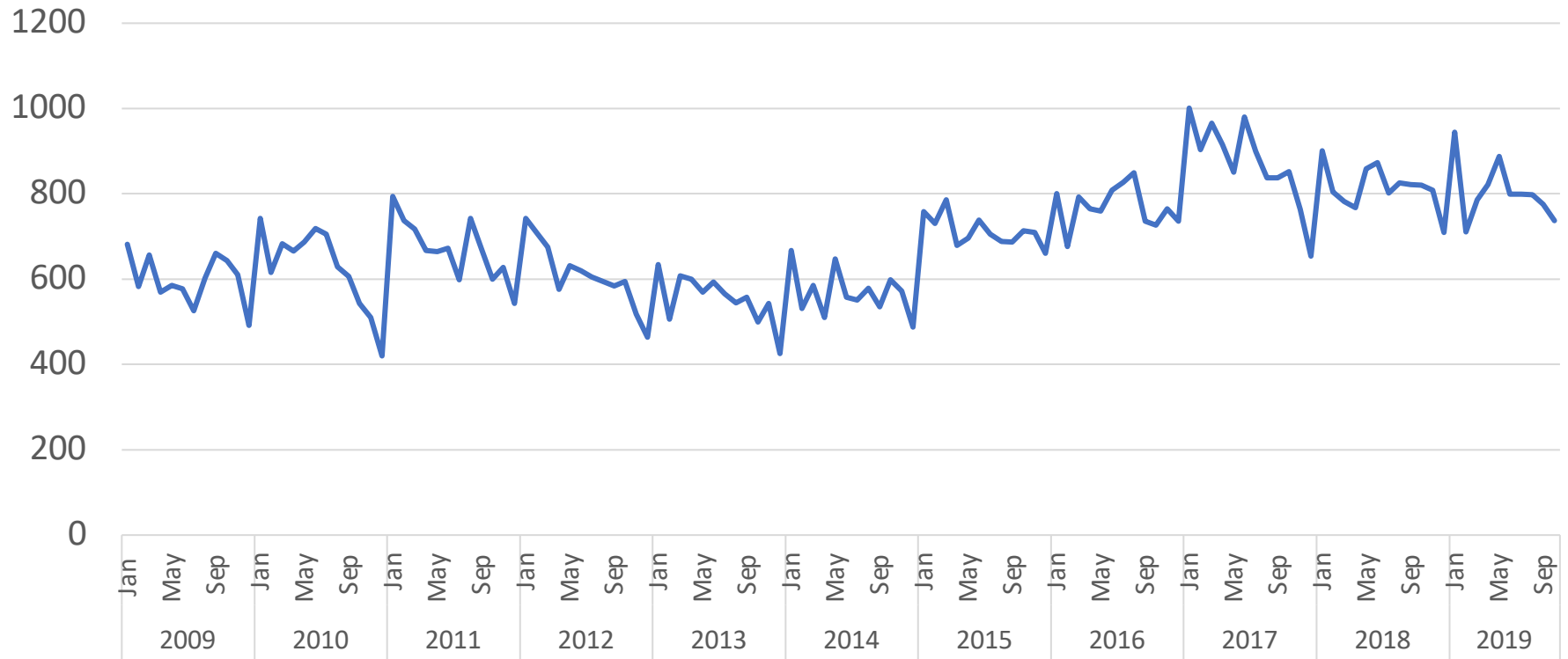
- **Unidimensional** on panel on a **group of individuals** (panel members) through **multiple time points** (ดูติดตามกลุ่มอย่างต่อเนื่องในหลายช่วงเวลา ไม่จำเป็นต้องเป็นคนเดิม)
- A time series is a collection of observations of well defined **data** items collected through **repeated measurements over equally spaced time intervals** (hourly, daily, weekly, monthly) (เก็บข้อมูลอย่างต่อเนื่องในหน่วยเวลาที่เท่าๆกัน เช่น ทุกชั่วโมง ทุกวัน ทุกสัปดาห์)

Data collected irregularly are not time series!!!

	Count
⊕ ม.ค.	111
⊕ ก.พ.	56
⊕ มี.ค.	69
⊕ เม.ย.	152
⊕ พ.ค.	340
⊕ มิ.ย.	508
⊕ ก.ค.	415
⊕ ส.ค.	282
⊕ ก.ย.	89

What are time series data?

ผู้ป่วยโรคปอดบวมที่มีอายุ 60 ปีขึ้นไป 2009-2019



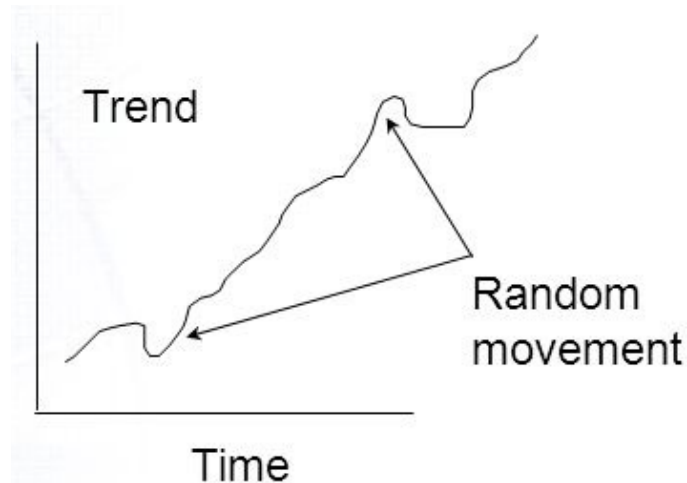


องค์ประกอบของอนุกรมเวลา

- A time series can have one or a **combination of the components** below. Using these components, we can project each component individually into the future, for example a future trend or future seasonality, and then combining them together to get a future forecast.
 - Trend (แนวโน้ม)
 - Seasonality (ฤดูกาล)
 - Cyclical (รอบ)
 - Error / Irregularity / Residual (ค่าความคลาดเคลื่อน)

องค์ประกอบของอนุกรมเวลา

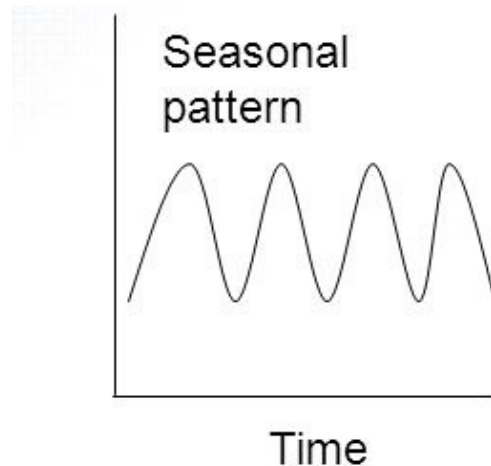
- **1) Trend (แนวโน้ม)**
 - Gradual upward/downward movement of data over time
 - Long term movement in a time series not related to calendar & irregular effects



องค์ประกอบของอนุกรมเวลา

• 2) Seasonality (ฤดูกาล)

- **Pattern of demand fluctuation** above or below the trend line that occurs **regularly**
- Short-term regular and repetitive variations which can be **as long as a year or as short as a few seconds**

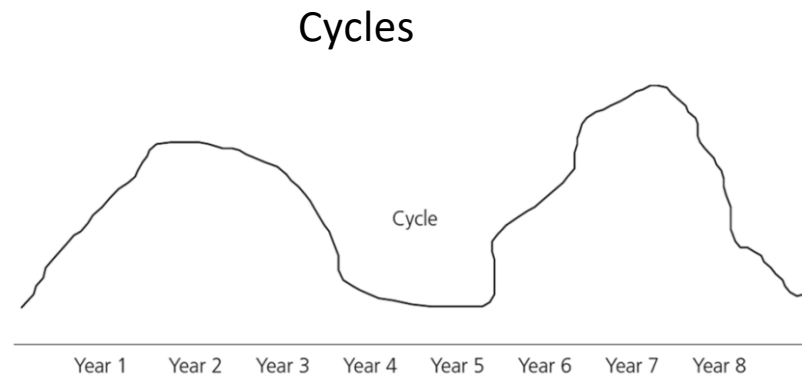




องค์ประกอบของอนุกรมเวลา

• 3) Cycles (รอบ)

- Pattern in the data that **occurs every several years** (usually tied to business cycle)
- Has a duration of **at least one year**
- Longer term thus requires many years of data to determine its repetitiveness or unusual circumstances

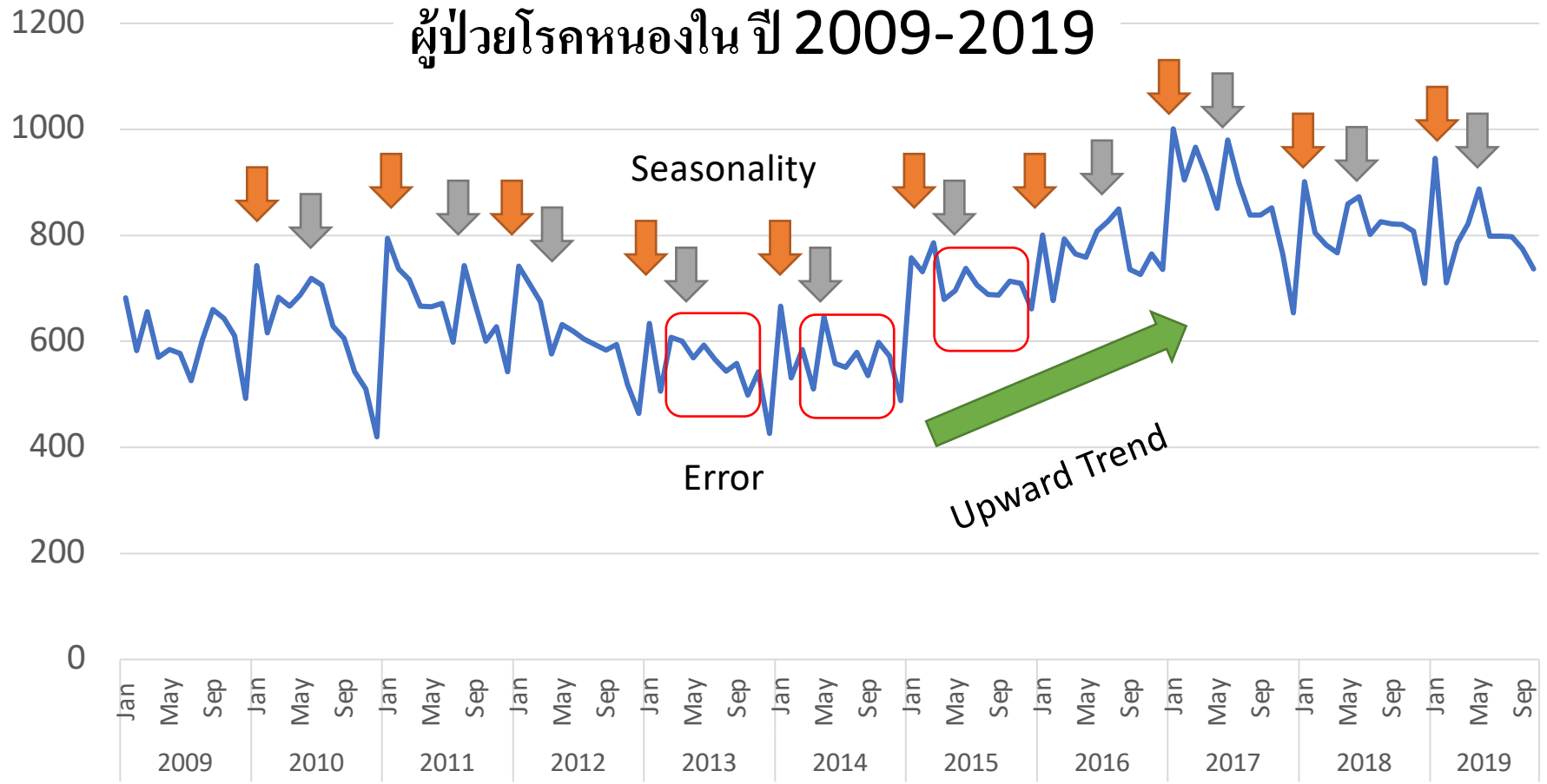


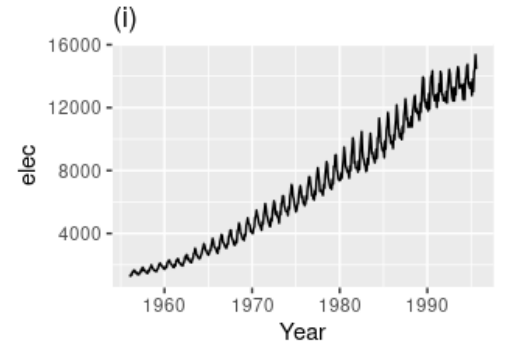
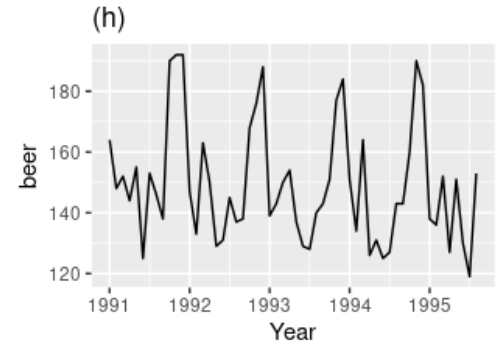
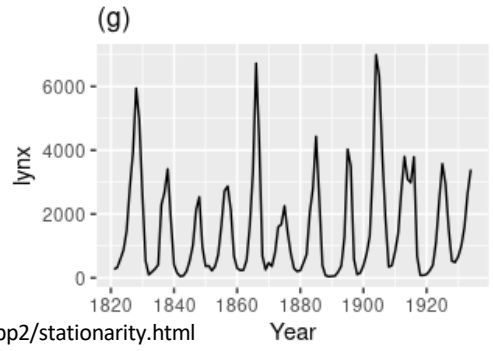
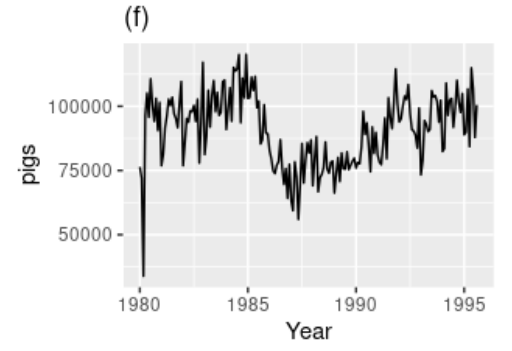
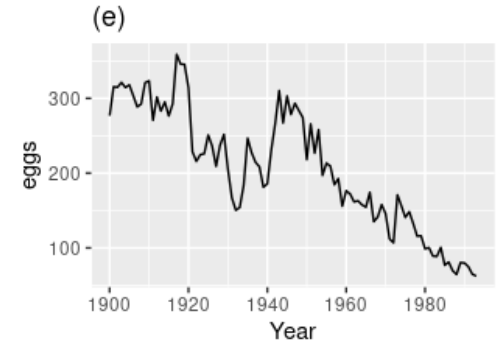
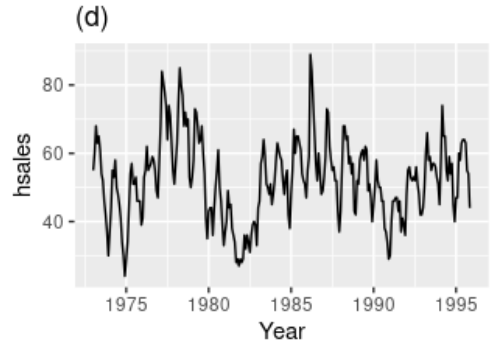
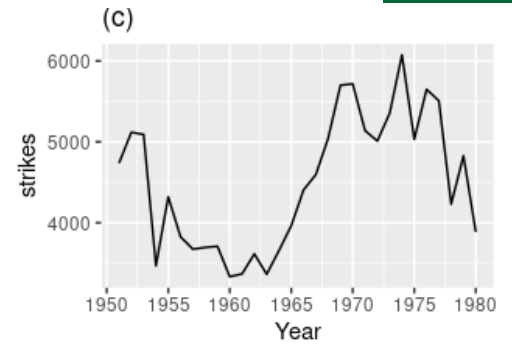
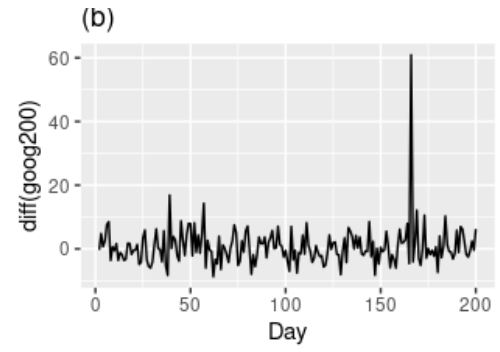
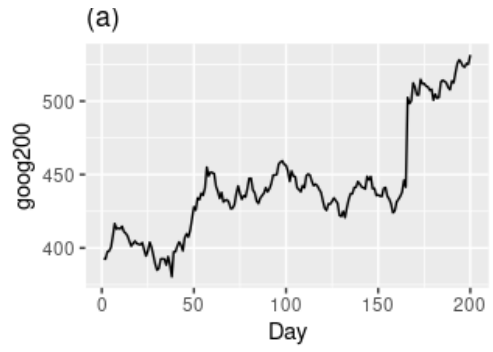
องค์ประกอบของอนุกรมเวลา

- **4) Error / Irregularity / Residual** (ค่าความคลาดเคลื่อน)
 - Short term **fluctuations** in the series which are **neither systematic nor predictable**
 - Appear as small random ups and downs and can dominate movements, which will mask the trend and seasonality

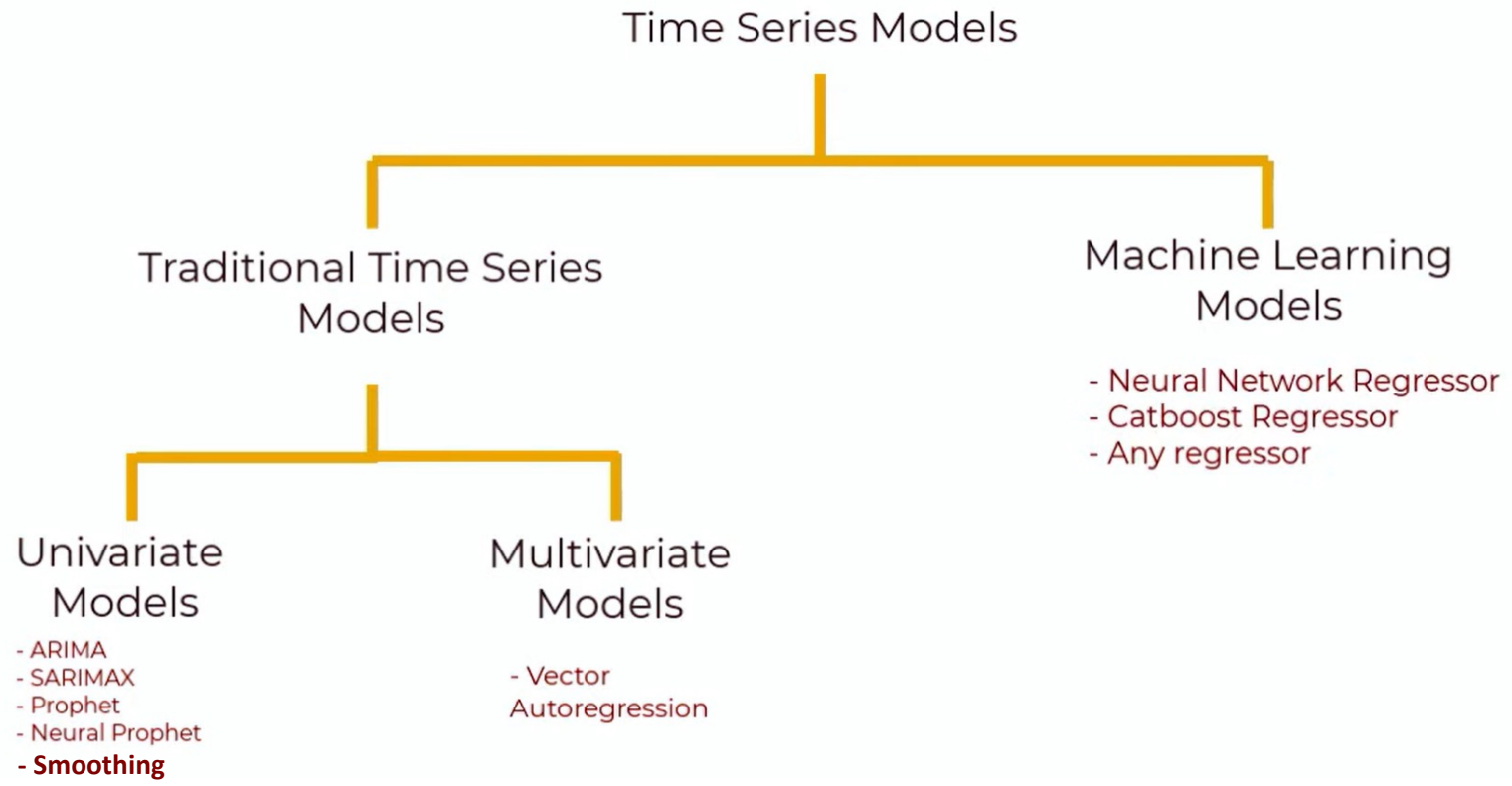


องค์ประกอบของอนุกรมเวลา





วิธีการพยากรณ์แบบอนุกรมเวลา





Model Selection

1. Smoothing
 1. Simple Moving Average (SMA)
 2. Single Exponential Smoothing (SES)
 3. Double Exponential Smoothing (DES)
 4. Triple Exponential Smoothing (TES; Additive model, Multiplicative model)
2. Autoregressive Integrated Moving Average (ARIMA)
 1. Autoregressive (AR)
 2. Moving Average (MA)
 3. Autoregressive Moving Average (ARMA)
 4. Autoregressive Integrated Moving Average (ARIMA)
 5. Seasonal Autoregressive Integrated Moving Average (SARIMAX)
3. Prophet
4. Vector Autoregression
5. Machine Learning Model

Method 1: Smoothing

Smoothing techniques are **kinds of data preprocessing techniques to remove noise from a data set**. This allows important patterns to stand out.



Method 1: Smoothing

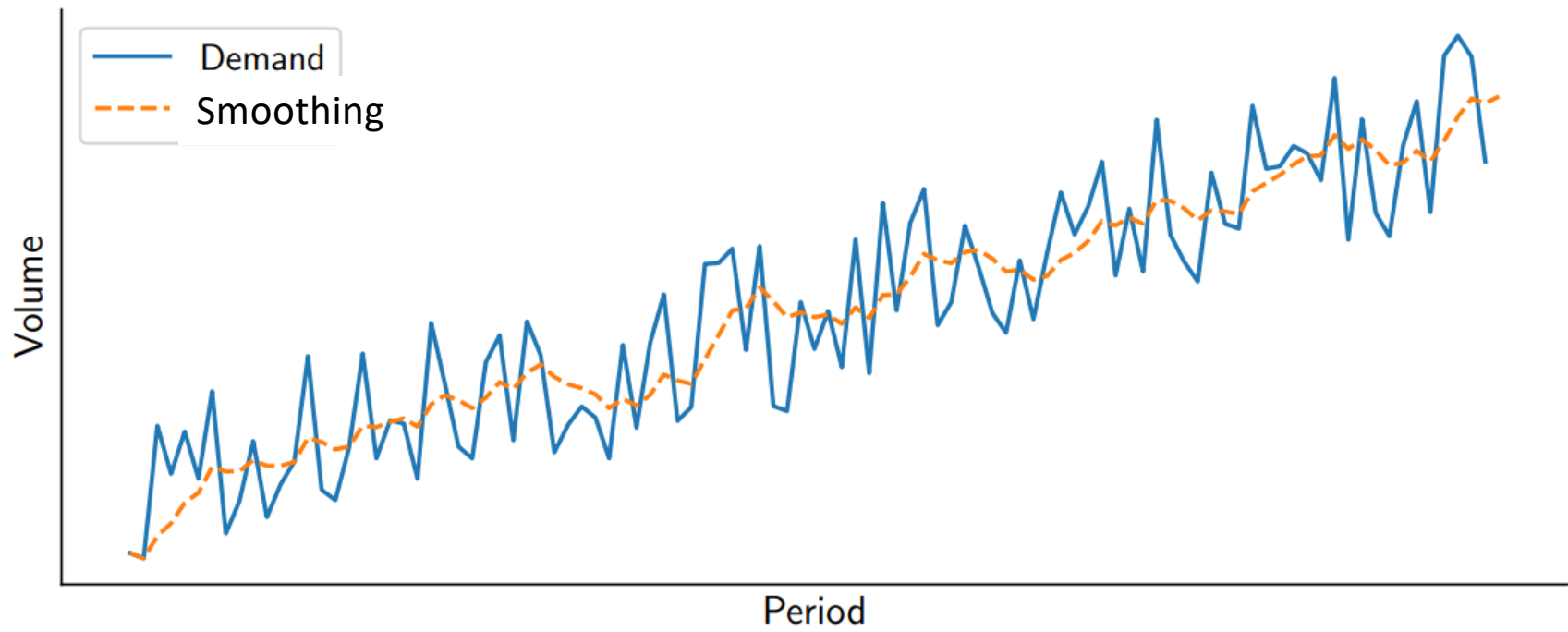


Figure 3.1: Demand level



Method 1.1: Simple Moving Average (เส้นค่าเฉลี่ย)

- Moving average is computed by taking the **average** of **several sequential data points**. (หาค่าเฉลี่ยจากข้อมูลย้อนหลัง)
- It is suitable when there is **no trend and no seasonality**, and forecast can be estimated using a **smoothing technique** and the **randomness is smoothed out**. (ใช้ในการจัดการค่าความคาดเคลื่อน)
- The basic principle is that demand observations that are **close to one another** are also likely to be **similar**
- Moving average model creates a new average as each new actual demand becomes available by dropping the oldest demand data, thus the selection of the sequential data is a **moving selection**.



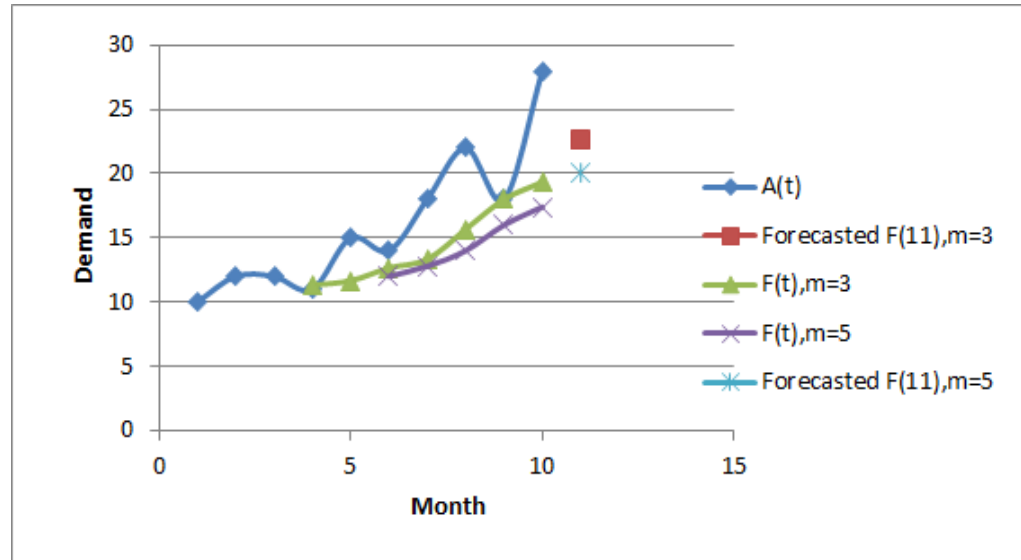
Method 1.1: Simple Moving Average (เส้นค่าเฉลี่ย)

Week	No of cases	Three-week moving average (m=3)
1	8	
2	10	
3	9	
4	11	$(8+10+9)/3 = 9$
5	10	$(10+9+11)/3 = 10$
6	13	$(9+11+10)/3 = 10$
7	?	$(11+10+13)/3 = 11.3$

Method 1.1: Simple Moving Average (เส้นค่าเฉลี่ย)

- Use different m to forecast demand for 11th month

Month	Demand A(t)	m = 3	m = 5
1	10		
2	12		
3	12		
4	11	11.33	
5	15	11.67	
6	14	12.67	12.00
7	18	13.33	12.80
8	22	15.67	14.00
9	18	18.00	16.00
10	28	19.33	17.40
11		22.67	20.00



$$F(t+1) = L(t) = \frac{\sum_{i=t-m+1}^t A(i)}{m}$$

$$m = 3, \quad F(4) = L(3) = \frac{10+12+12}{3} = 11.33$$

$$m = 5, \quad F(6) = L(5) = \frac{10+12+12+11+15}{5} = 12.0$$

Method 1.2 – Single Exponential Smoothing

- Single Exponential Smoothing is very much an exponential moving average, which means that there will be **no trend** component in the data.

Need an
initial estimate

$$F(t+1) = F(t) + \alpha(A(t) - F(t)) \quad 0 \leq \alpha \leq 1$$

- $F(t)$ = Forecast at time t
 - $F(t+1)$ = Forecast at time $t+1$
 - $A(t)$ = Actual at time t
 - $\alpha = 1$, next forecast is sensitive to past forecast error
 - $\alpha = 0$, next forecast is exactly same as previous forecast
- We can see that the next forecast is equals to the previous forecast adjusted by past **forecast error**, given by $\alpha(A(t) - F(t))$

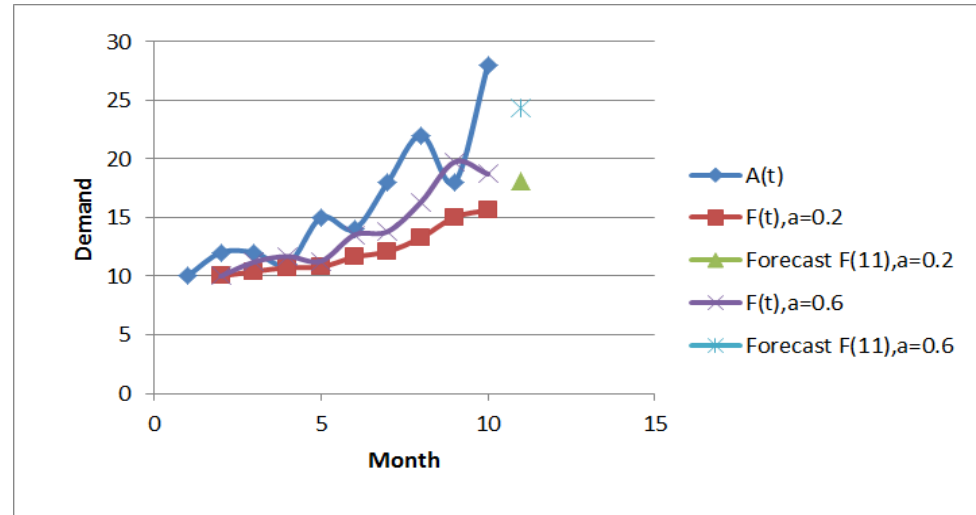
Method 2 – Single Exponential Smoothing

- Forecast demand for 11th month

		$\alpha = 0.2$	$\alpha = 0.6$
Month	Demand A(t)	Forecast F(t)	Forecast F(t)
1	10	10.00	10.00
2	12	10.00	10.00
3	12	10.40	11.20
4	11	10.72	11.68
5	15	10.78	11.27
6	14	11.62	13.51
7	18	12.10	13.80
8	22	13.28	16.32
9	18	15.02	19.73
10	28	15.62	18.69
11		18.09	24.28

We assume an initial value for F(1) or let F(1) = average of actual demand, A(1) to A(10)

$$F(t+1) = F(t) + \alpha(A(t) - F(t))$$



$$F(1) = 10,$$

$$\alpha = 0.2,$$

$$\alpha = 0.6,$$

$$F(2) = \alpha A(1) + (1 - \alpha)F(1)$$

$$F(3) = \alpha A(2) + (1 - \alpha)F(2)$$

$$F(2) = 0.2 * 10 + 0.8 * 10 = 10$$

$$F(3) = 0.2 * 12 + 0.8 * 10 = 10.4$$

$$F(2) = 0.6 * 10 + 0.4 * 10 = 10$$

$$F(3) = 0.6 * 12 + 0.4 * 10 = 11.2$$



Model Selection

1. Smoothing
 1. Simple Moving Average (SMA)
 2. Single Exponential Smoothing (SES)
 3. Double Exponential Smoothing (DES)
 4. Triple Exponential Smoothing (TES; Additive model, Multiplicative model)
2. Autoregressive Integrated Moving Average (ARIMA)
 1. Autoregressive (AR)
 2. Moving Average (MA)
 3. Autoregressive Moving Average (ARMA)
 4. Autoregressive Integrated Moving Average (ARIMA)
 5. Seasonal Autoregressive Integrated Moving Average (SARIMAX)
3. Prophet
4. Vector Autoregression
5. Machine Learning Model

Autoregressive (AR) & Moving Average (MA) models

- **AR:** The term *autoregression* indicates that it is a regression of the ***variable against itself***. (การพยากรณ์สนใจในค่าของตัวเองในอดีต)

$$\bar{y}_t = \alpha_{t-p}y_{t-p} + \alpha_{t-p+1}y_{t-p+1} + \dots + \alpha_{t-1}y_{t-1}$$

- **MA:** Rather than using past values of the forecast variable in a regression, a moving average model uses ***past forecast errors*** in a regression-like model. (พยากรณ์โดยใช้ค่าความคาดเคลื่อนในอดีต)

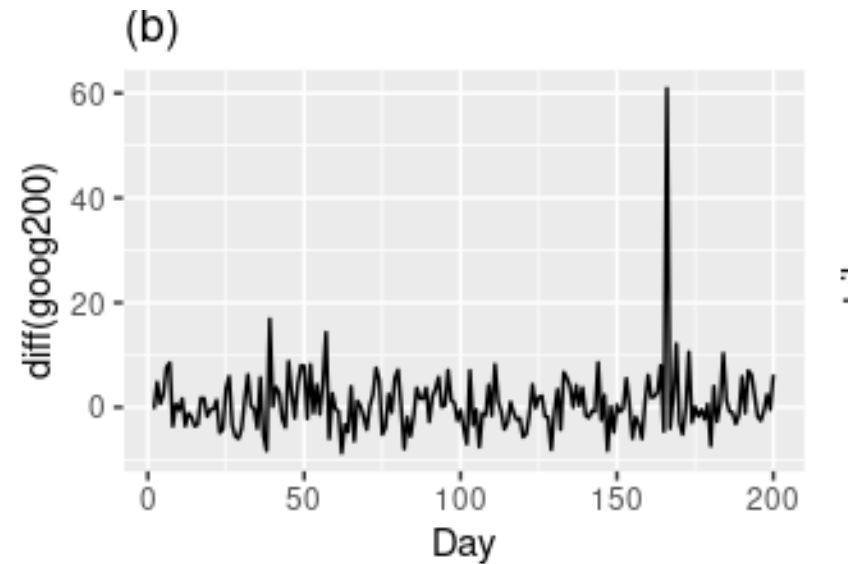
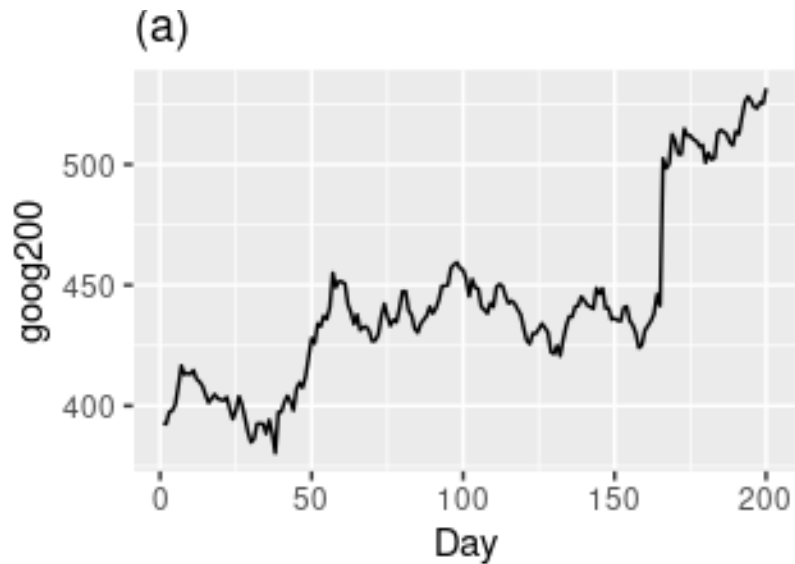
$$\bar{y}_t = \theta_{t-q} \varepsilon_{t-q} + \theta_{t-q+1} \varepsilon_{t-q+1} + \dots + \theta_{t-1} \varepsilon_{t-1}$$

Autoregressive Integrated Moving Average (ARIMA)

- Before we introduce ARIMA models, we must first discuss
 - the concept of *stationarity* -> *No trend, No Seasonality*
 - the technique of *differencing* time series.
- Many non-stationary time series can be converted to a stationary time series by taking d^{th} order *differences*.
- Included Differencing (I) into ARMA. (เพิ่ม Differencing ใน ARMA เพื่อจัดจากอนุกรมเวลาไม่คงที่)



Differencing



Prophet Model

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t$$

- **$g(t)$** : It represents the ***trend*** and the objective is to capture the general trend of the series.
- **$s(t)$** : It is the ***Seasonality*** component. The number of advertisement views might also depend on the season.
- **$h(t)$** : The ***Holidays*** component. We use the information for holidays which have a clear impact on most business time series. Note that holidays vary between years, countries, etc. and therefore the information needs to be explicitly provided to the model.
- The **error term** ε_t stands for random fluctuations that cannot be explained by the model. As usual, it is assumed that ε_t follows a normal distribution $N(0, \sigma^2)$ with zero mean and unknown variance σ that has to be derived from the data.

Multivariate Forecasting



PM2.5

- ปัจจัยด้านสิ่งแวดล้อมที่มีต่อการเกิดโรคไข้เลือดออก
- อุณหภูมิ
 - ความชื้นสัมพัทธ์
 - ปริมาณน้ำฝน
 - ค่าดัชนี HI/CI
 - จำนวนครั้งการระบาด (Cycles)

Build up
Prediction Model
By Machine
Learning

การระบาดไข้เลือดออก

Vector Autoregression (VAR)

- In a VAR model, each variable is a linear function of the **past values of itself** and the **past values of all the other variables**.

Variable y1	Variable y2	Variable y1	Variable y2
$y1_{t-n}$	$y2_{t-n}$	$y1_{t-n}$	$y2_{t-n}$
...
$Y1_{t-2}$	$Y2_{t-2}$	$Y1_{t-2}$	$Y2_{t-2}$
$Y1_{t-1}$	$Y2_{t-1}$	$Y1_{t-1}$	$Y2_{t-1}$
$y1_t$	$y2_t$	$y1_t$	$y2_t$

$$\begin{bmatrix} y1(t) \\ y2(t) \end{bmatrix} = \begin{bmatrix} a1 \\ a2 \end{bmatrix} + \begin{bmatrix} w11 & w12 \\ w21 & w22 \end{bmatrix} \begin{bmatrix} y1(t-1) \\ y2(t-1) \end{bmatrix} + \begin{bmatrix} e1(t) \\ e2(t) \end{bmatrix}$$

$$\begin{bmatrix} y1 \\ y2 \\ \vdots \\ yk \end{bmatrix} = \begin{bmatrix} a1 \\ a2 \\ \vdots \\ ak \end{bmatrix} + \begin{bmatrix} w11 & \dots & \\ w21 & \dots & \\ \vdots & \vdots & \\ wk1 & \dots & \end{bmatrix} \begin{bmatrix} y1(t-1) \\ y2(t-1) \\ \vdots \\ yk(t-1) \end{bmatrix} + \dots + \begin{bmatrix} w'11 & \dots & \\ w'21 & \dots & \\ \vdots & \vdots & \\ w'k1 & \dots & \end{bmatrix} \begin{bmatrix} y1(t-p) \\ y2(t-p) \\ \vdots \\ yk(t-p) \end{bmatrix} + \begin{bmatrix} e1 \\ e2 \\ \vdots \\ ek \end{bmatrix}$$

K x 1

K x 1

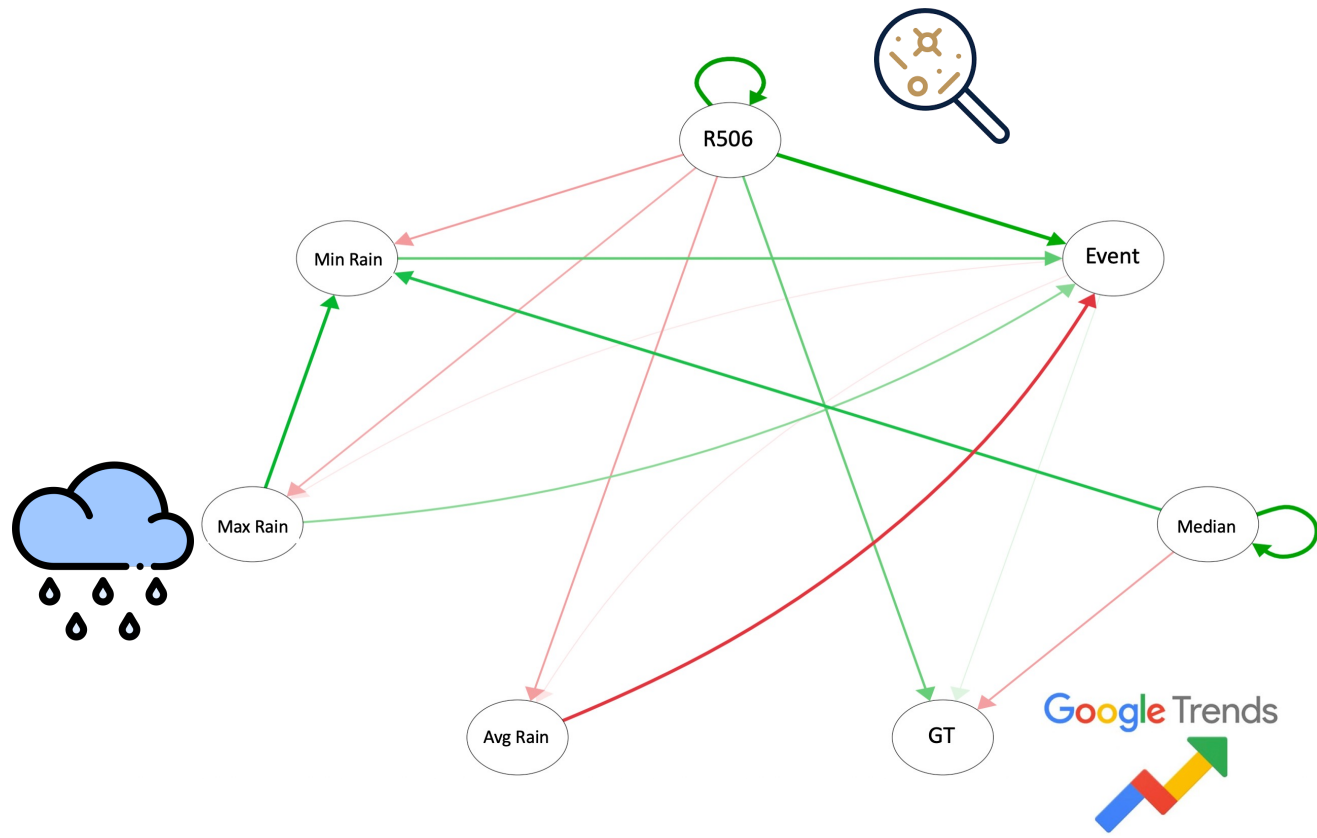
K x K

K x 1

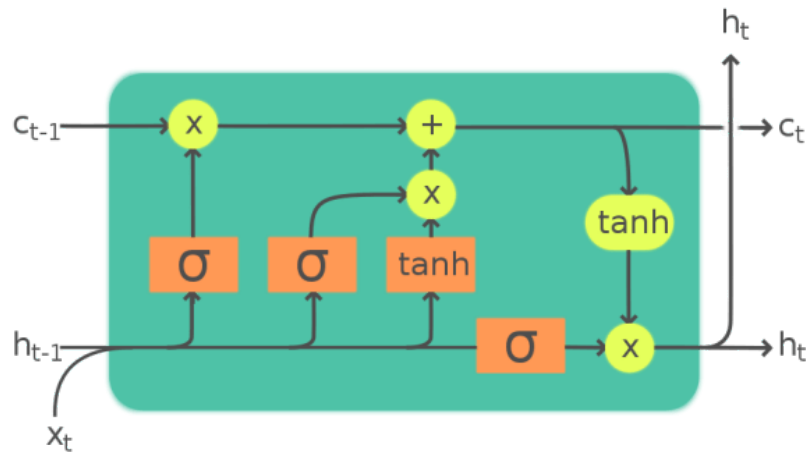
K x K

K x 1

Vector Autoregression (VAR)



Machine Learning: Long short-term memory Neural Network



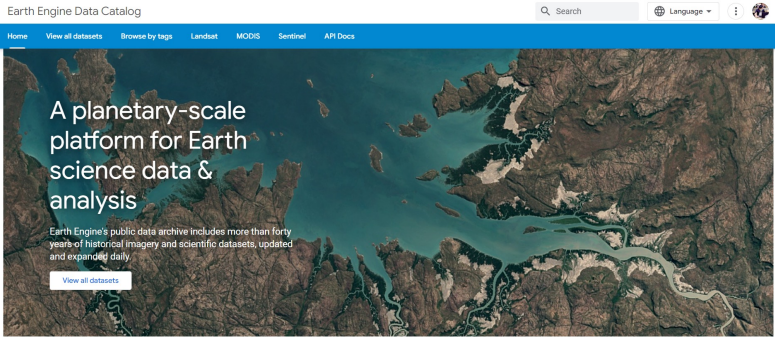
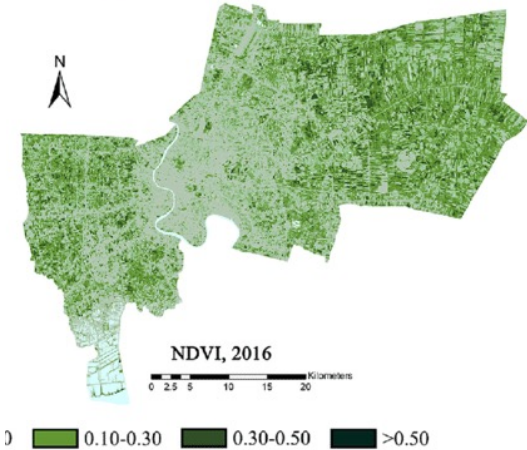
Legend:

Layer	ComponentwiseCopy	Concatenate

Time Series Forecasting of Dengue Fever Using Google Earth Engine & Machine Learning Algorithms

Case Study : Kang Krachan District, Phetchaburi

Nattanya MEEWUTTISOM, Suphanat Wongsanuphat



Climate and Weather

Land Surface Temperature (LST)

Precipitation , Numbers of Rainy Days

Humidity

Normalized Different Vegetation Index (NDVI)

CASE	MODEL	N_SPLITS	BEST PARAMETERS			BEST_MAPE
			BATCH_SIZE	EPOCHS		
case_n	LSTM_day_rained1	5	64	150	64.21%	
case_n	LSTM_precipitation_1	5	64	150	61.56%	
case_n	LSTM_HI	5	64	150	67.19%	
case_n	LSTM_CI	5	64	150	66.85%	
case_n	LSTM_BI	5	128	150	65.37%	
case_n	LSTM_humidity	5	128	150	63.54%	
case_n	LSTM_day_rained1	5	64	150	64.21%	
case_n	LSTM_[day_rained1,precipitation_1]	5	64	100	49.82%	
case_n	LSTM_[day_rained1,precipitation_1,HI]	5	128	100	52.77%	
case_n	LSTM_[day_rained1,precipitation_1,CI]	5	32	100	54.44%	
case_n	LSTM_[day_rained1,precipitation_1,BI]	5	64	100	49.15%	
case_n	LSTM_[day_rained1,precipitation_1,humidity]	5	64	100	53.99%	
case_n	LSTM_[day_rained1,precipitation_1,HI,CI]	5	32	50	44.69%	
case_n	LSTM_[day_rained1,precipitation_1,HI,BI]	5	128	50	43.01%	
case_n	LSTM_[day_rained1,precipitation_1,HI,humidity]	5	128	50	43.89%	
case_n	LSTM_[day_rained1,precipitation_1,HI,CI,BI]	5	64	50	47.49%	
case_n	LSTM_[day_rained1,precipitation_1,HI,CI,BI,humidity]	5	32	150	52.22%	
case_n	LSTM_[day_rained1,precipitation_1,HI,CI,humidity]	5	128	50	50.25%	
case_n	LSTM_[day_rained1,precipitation_1,HI,BI,humidity]	5	32	50	47.12%	
case_n	LSTM_[day_rained1,precipitation_1,BI,humidity]	5	32	50	40.77%	
case_n	LSTM_[day_rained1,precipitation_1,CI,BI]	5	128	50	40.84%	
case_n	LSTM_[day_rained1,precipitation_1,CI,humidity]	5	32	50	42.90%	
case_n	LSTM_[day_rained1,HI]	5	128	100	48.37%	
case_n	LSTM_[day_rained1,HI,CI]	5	128	100	57.51%	
case_n	LSTM_[day_rained1,HI,BI]	5	64	100	55.50%	
case_n	LSTM_[day_rained1,HI,humidity]	5	32	100	57.31%	
case_n	LSTM_[day_rained1,HI,CI,BI]	5	64	50	48.20%	

Select Methods

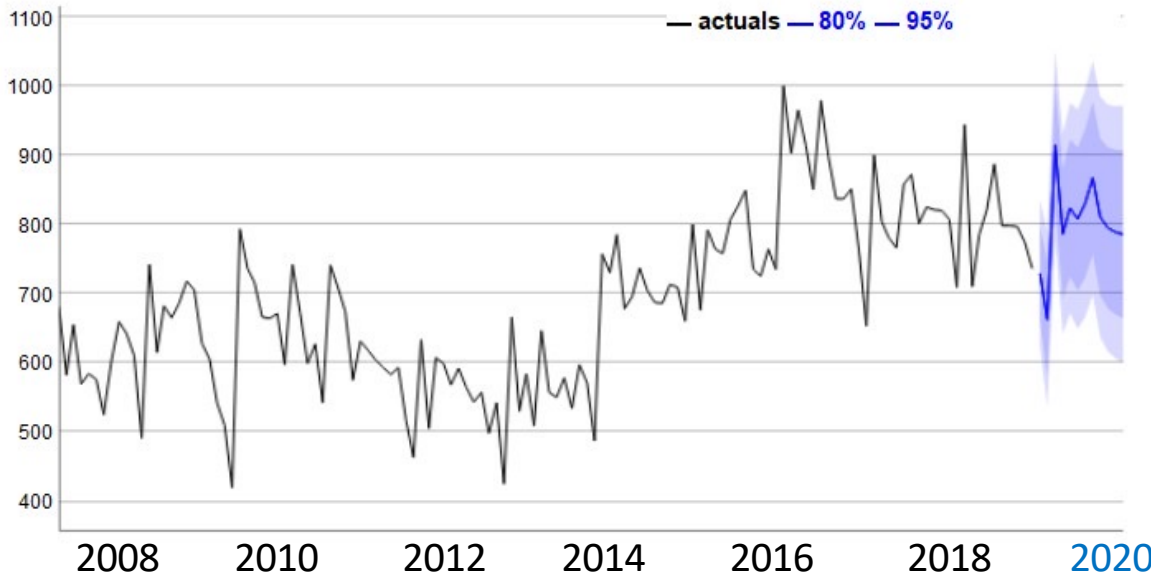
1. Select the forecasting method to use by first **plotting** out the past data to see if there is trend and seasonality. (Plot กราฟ)
2. For each **appropriate method**, determine the **optimized parameters** by minimizing SSE. (เลือกวิธีการ และ **parameters** ที่เหมาะสม)
3. Compare selected methods using quantitative measures. (เปรียบเทียบความแม่นยำของแต่ละวิธีการ)

Method	Smoothing Function	Trend (T)	Seasonality (S)	Holiday Period	Multivariable	Parameters (optimized)
Moving Average	✓					m
Single ES	✓					α
Double ES	✓	✓				α, β
Triple ES (Add)	✓	✓	✓			α, β, γ
Triple ES (Mult)	✓	✓	✓			α, β, γ
ARIMA	✓	✓				p,d,q
SARIMA	✓	✓	✓			p,d,q P,D,Q
Prophet	✓	✓	✓	✓		
VAR	✓	✓	✓		✓	
Machine Learning	✓	✓	✓	✓	✓	

Time Series Forecasting

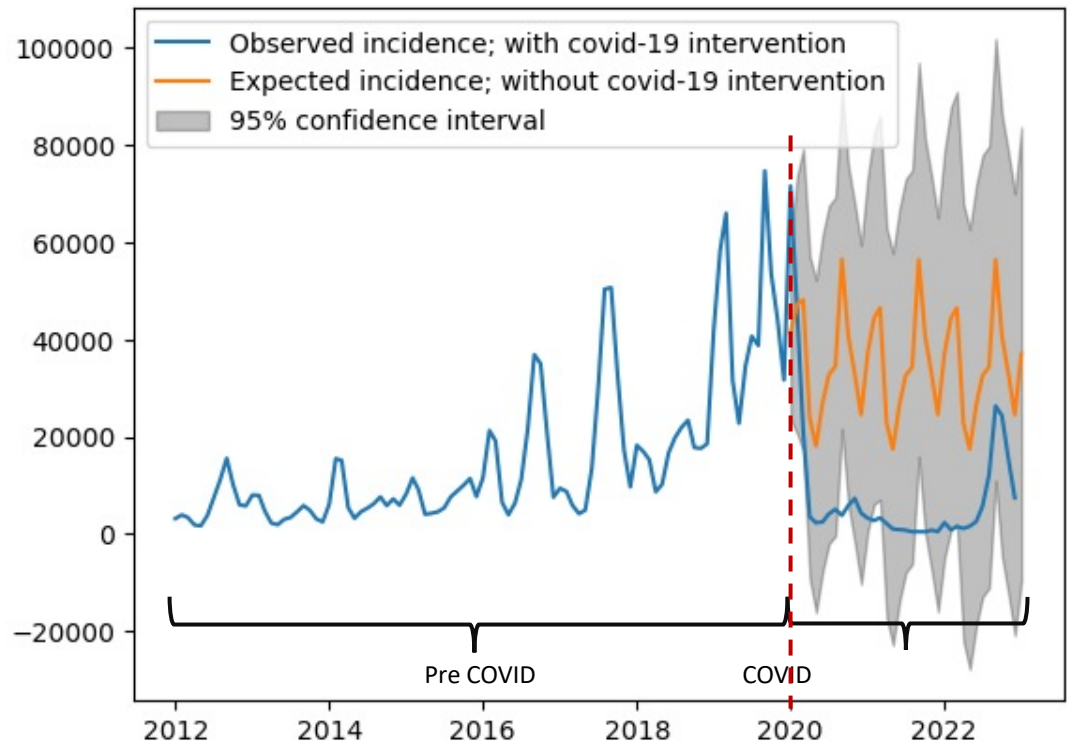
ทำนายอนาคต: จำนวนผู้ป่วย **GC** ในปี **2020** จะเป็นเท่าไร?

Number of GC cases during 2008-2019 and prediction in 2020



Time Series Forecasting

ประเมินผลมาตรการ: ผลมาตรการควบคุมโรคCOVID ส่งผลต่อโรคไข้หวัดใหญ่อย่างไร?

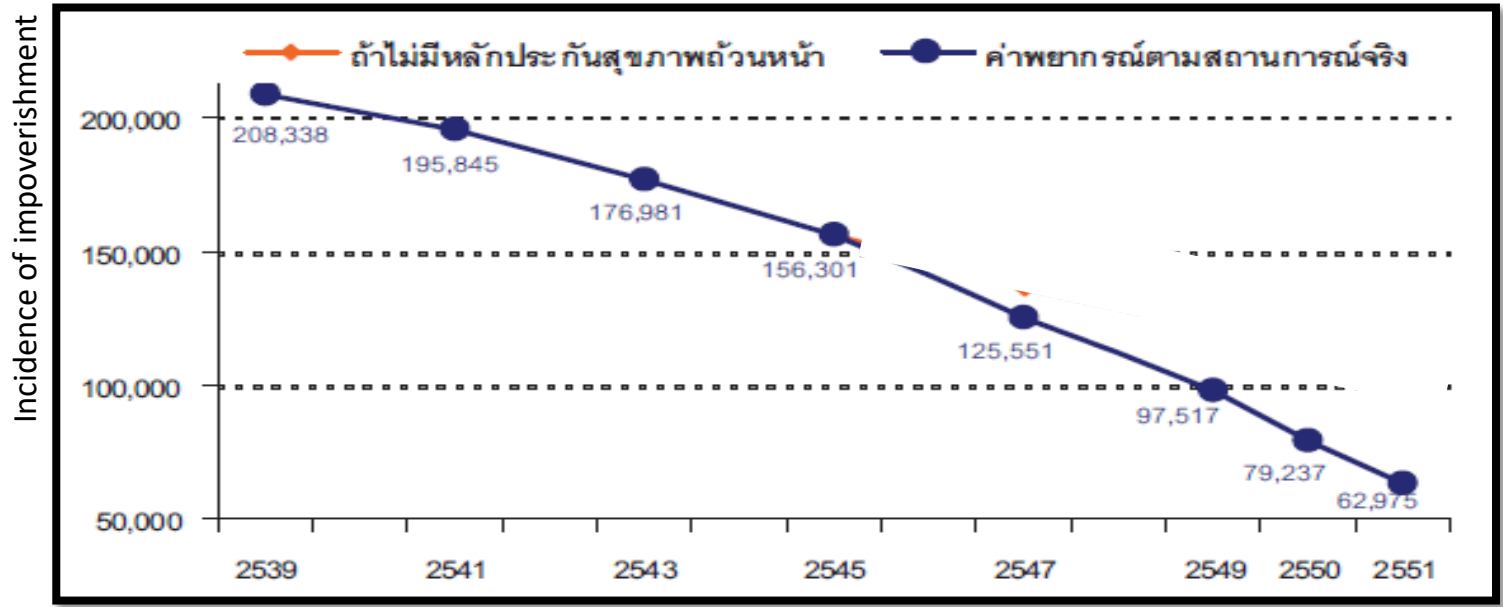


Source: API COVID, DDC

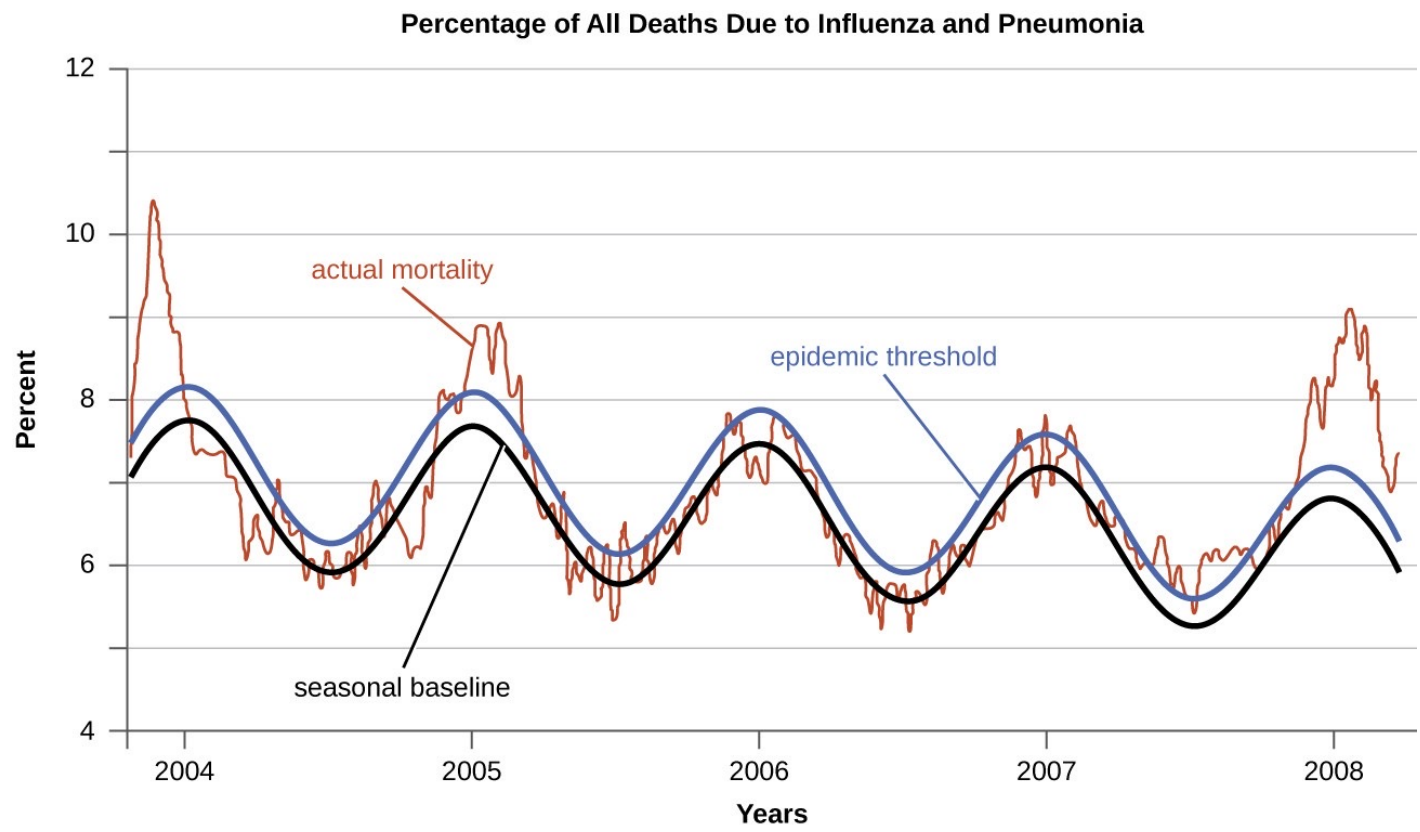
Time Series Forecasting

What if Thailand government didn't provide Universal Health coverage (บัตรทอง) since 2002?

Impact of UC on incidence of impoverishment in Thailand



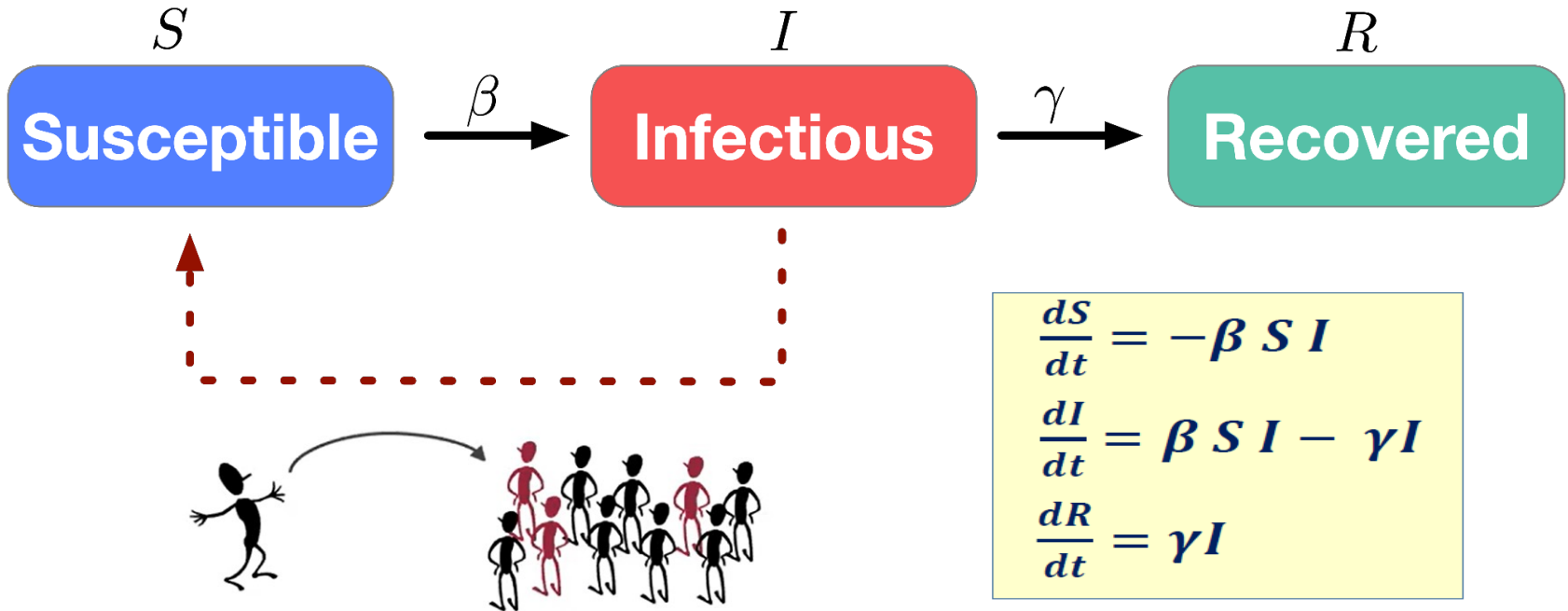
แจ้งเตือนการระบาด (Early Warning System)



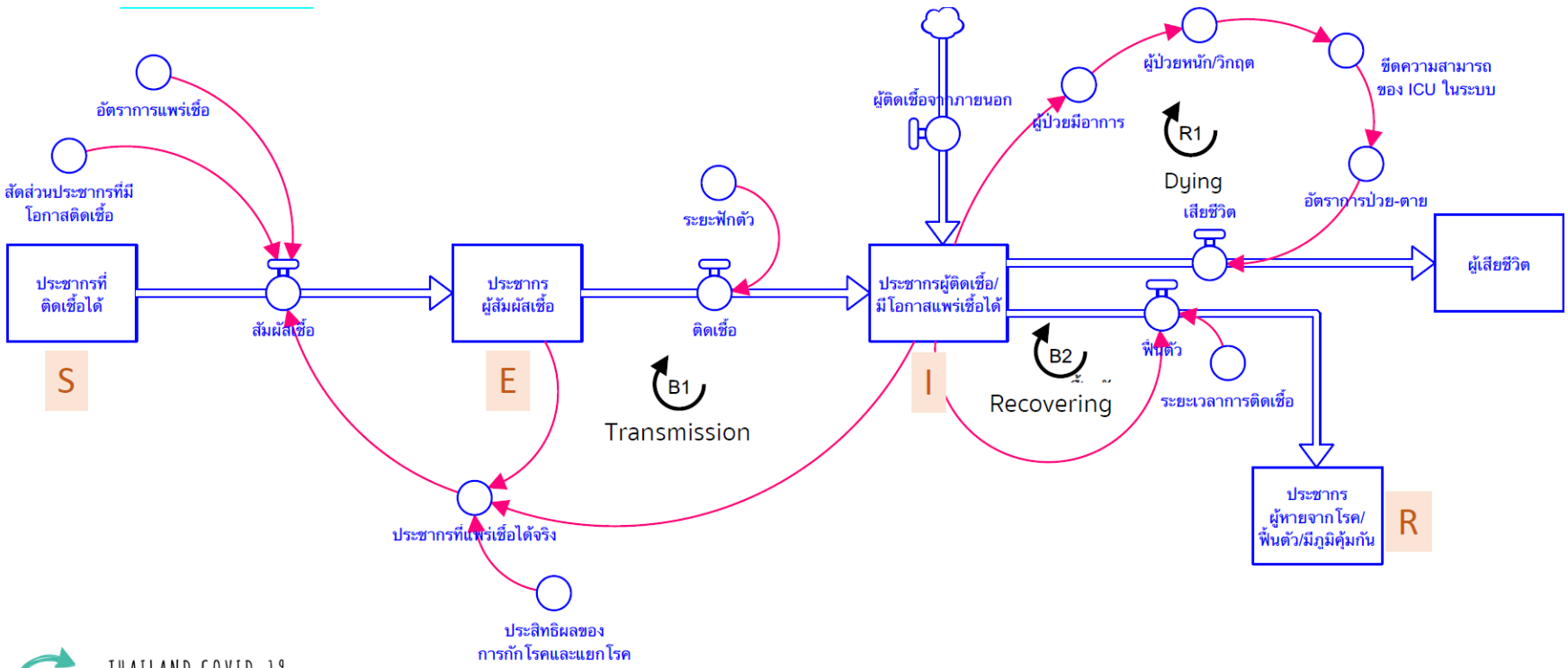


Compartmental Model

Compartmental Modeling

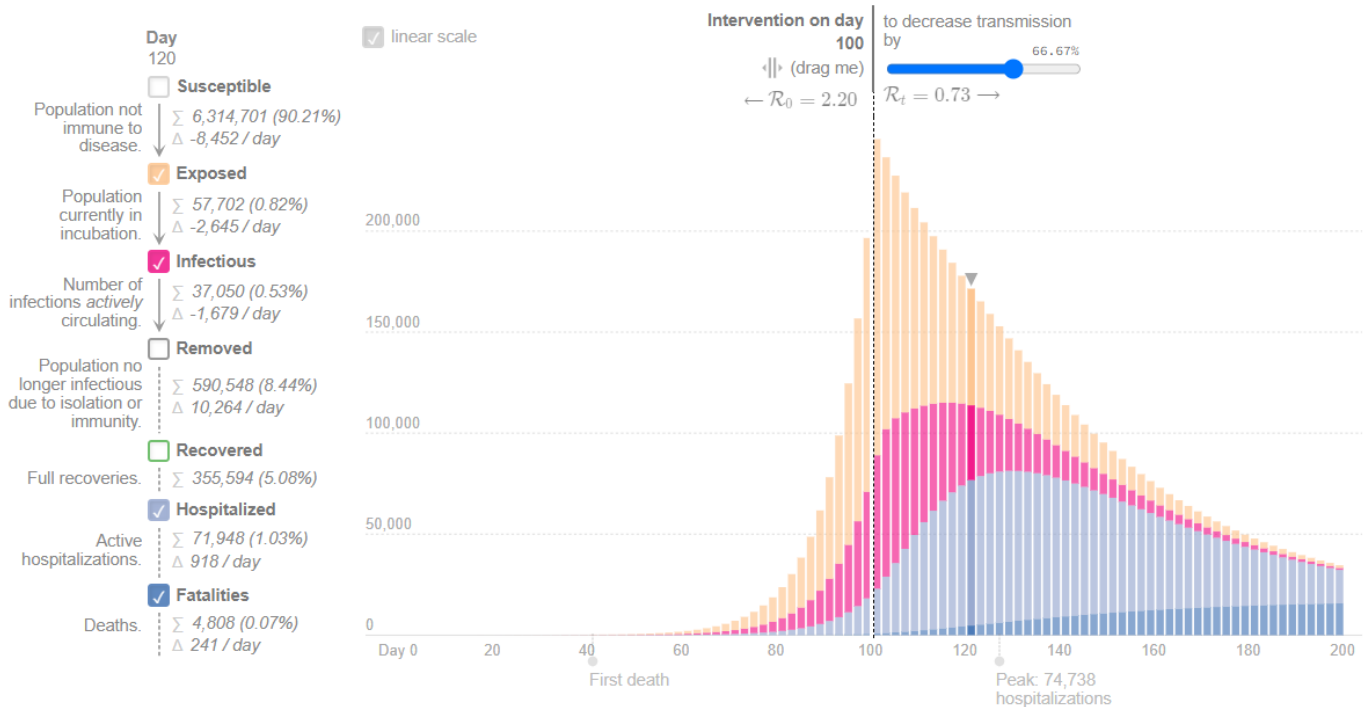


Compartmental Modeling



Epidemic Calculator

<https://gabgoh.github.io/COVID/index.html>



Transmission Dynamics

Population Inputs Size of population. $7,000,000$ Number of initial	Basic Reproduction Number \mathcal{R}_0 Measure of contagiousness: the number of secondary infections each infected individual produces.	Transmission Times Length of incubation period, T_{inc} . 5.20 days Duration patient is
---	---	---

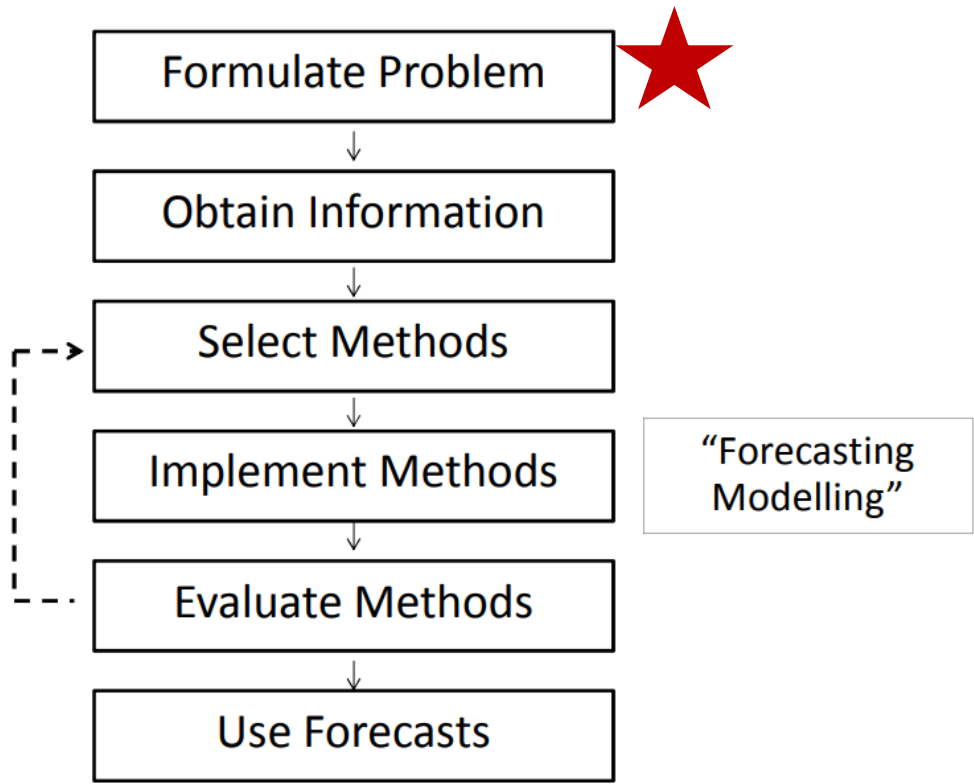
Clinical Dynamics

Mortality Statistics Case fatality rate. 2.00% Time from end of	Recovery Times Length of hospital stay. 28.6 Days Recovery time for mild	Care statistics Hospitalization rate. 20.00% Time to hospitalization
--	--	---



ขั้นตอนการพยากรณ์โรค

ขั้นตอนการพยากรณ์โรค

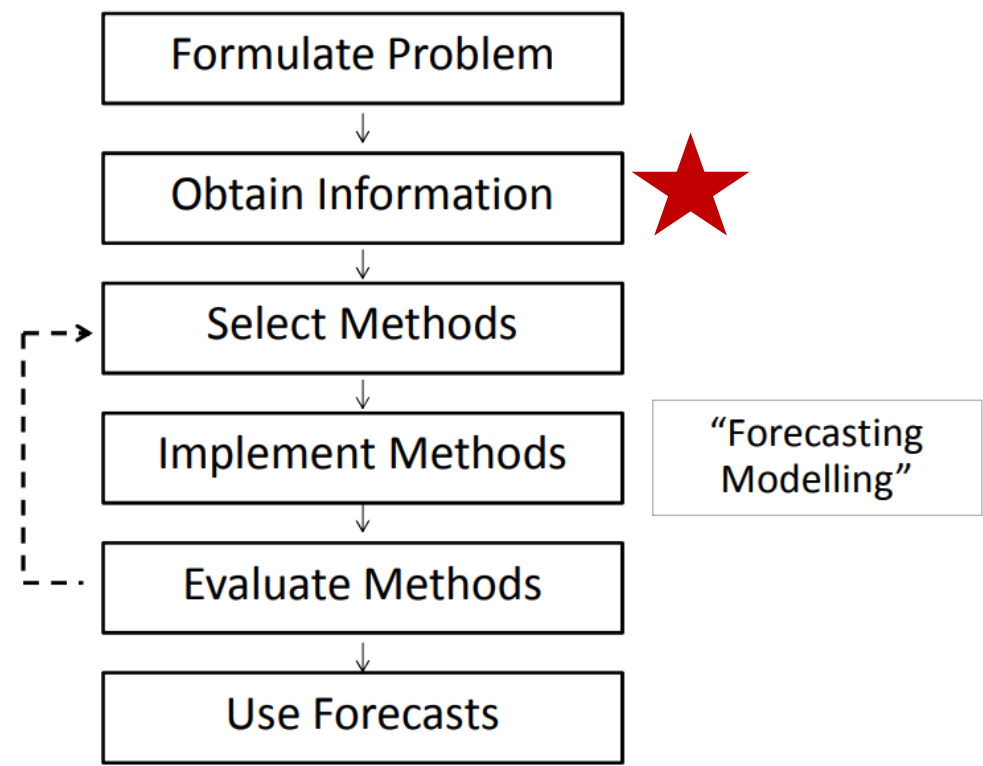


1. ตั้งคำถามการพยากรณ์

จำนวนสูงอายุที่ป่วยเป็นปอดอักเสบจะเป็นเท่าไรในปี **2020?**



ขั้นตอนการพยากรณ์โรค



2.รวบรวมข้อมูล

R506 Surveillance

- เป็นระบบที่ได้รับความร่วมมือจากสำนักงานสาธารณสุขจังหวัด โรงพยาบาล และสถานอนามัยทุกแห่ง (โรงพยาบาลรัฐทุกแห่ง โรงพยาบาลเอกชนยังไม่ครอบคลุมทั้งหมด) ในการเฝ้าระวังโรค/ภัย ที่อาจเกิดการระบาดได้ โดยมีวัตถุประสงค์เพื่อการป้องกันควบคุมโรค/ภัย เป็นหลัก มิใช่เป็นรายงานสถิติของโรคนี้ๆ
- ส่วนใหญ่ ใช้นิยามผู้ป่วยเป็น "ผู้ป่วยที่สงสัย(suspect)" ไม่ใช่ "ผู้ป่วยที่ยืนยัน(confirm)"

- มีโรคต้องรายงาน 57 โรค

AEFI	Food_Poisoning	Meningitis,uns	Typhoid
AFP	Genital Molluscum Contagiosum	Meningococcal Meningitis	Vaginal trichomoniasis
Anthrax	H.conjunctivitis	Mumps	Zika virus
Avian Influenza	HepatitisA	Paratyphoid	
Botulism	HepatitisB	Pediculosis Pubis	
Brucellosis	HepatitisC	Pertussis	
Chickenpox	HepatitisD	Pneumonia	
Chikungunya	HepatitisE	Poliomyelitis	
Cholera	Hepatitis_uns	Pulmonary T.B	
D.H.F	HFM	PUO	
D.H.Fshock syndrome	Influenza	Rabies	
Dengue fever	Japanese B encephalitis	Rubella	
Diarrhoea	Kala azar	Scarlet fever	
Diphtheria	Leprosy	Scrub Typhus	
Dysentery,Amoebic	Leptospirosis	Streptococcus suis	
Dysentery,Bacillary	Liver fluke	T.B. other organs	
Encephalitis uns	Malaria	T.B.meningitis	
Enterovirus Fever	Measles	Tetanus exc.Neo	
Eosinophilic Meningitis	Measles c Complication	Tetanus neonatorum	
Filariasis	Melioidosis	Trichinosis	



Is this dataset time series data?

disease	sex	agey	agem	aged	marietal	race	race1	occupat	addrcode	metropol	hospital	type	result	hserv	datesick	datedefin	datereach	datedeath	
30	1	24	5	24	1	1	1	0	3 85030404	2	3	3	1	3	850301	1/1/2022	1/1/2022	3/1/2022	0000-00-0
30	1	54	0	0	2	1	0	1 86060209	1	3	2	2	1	1	860601	1/1/2022	1/1/2022	#####	0000-00-0
30	1	63	5	13	2	1	0	1 85030406	2	3	3	1	3	3	850301	1/1/2022	1/1/2022	3/1/2022	0000-00-0
30	1	11	6	4	1	3	1	6 63080202	2	3	3	1	1	1	630801	1/1/2022	2/1/2022	5/1/2022	0000-00-0
30	1	27	10	10	2	1	0	3 85050105	1	3	3	1	1	1	850501	1/1/2022	2/1/2022	4/1/2022	0000-00-0
30	1	5	1	3	1	1	0	11 63080301	2	3	3	1	1	1	630801	1/1/2022	3/1/2022	5/1/2022	0000-00-0
30	1	4	4	0	1	1	0	11 63050403	2	3	3	1	3	3	630501	1/1/2022	3/1/2022	#####	0000-00-0
30	1	30	0	0	1	1	0	1 82070601	2	4	4	1	1	1	820701	1/1/2022	4/1/2022	5/1/2022	0000-00-0
30	1	24	0	0	1	1	0	1 82070601	2	4	4	1	1	1	820701	1/1/2022	4/1/2022	5/1/2022	0000-00-0
30	1	25	8	16	1	1	0	1 85050201	1	3	3	1	1	1	850501	1/1/2022	5/1/2022	7/1/2022	0000-00-0
30	1	16	0	8	1	3	1	1 63080409	1	3	3	1	1	1	630801	1/1/2022	6/1/2022	9/1/2022	0000-00-0
30	2	24	0	8	1	3	1	11 63080303	2	3	3	1	1	1	630801	1/1/2022	6/1/2022	9/1/2022	0000-00-0
30	1	28	0	0	2	1	0	10 86010208	1	3	3	2	1	1	860101	1/1/2022	6/1/2022	#####	0000-00-0
30	1	13	3	23	1	1	0	6 85050201	1	3	3	1	1	1	850501	1/1/2022	7/1/2022	9/1/2022	0000-00-0
30	1	15	1	0	1	1	0	6 63050302	2	3	3	1	3	3	630501	1/1/2022	8/1/2022	#####	0000-00-0
30	1	15	0	29	1	1	0	6 63080202	2	3	3	1	1	1	630801	1/1/2022	9/1/2022	#####	0000-00-0
30	2	8	0	0	1	1	0	6 95040307	2	3	3	1	3	3	950401	1/1/2022	9/1/2022	#####	0000-00-0
30	1	25	5	5	2	1	0	15 85030208	2	3	3	1	3	3	850301	1/1/2022	9/1/2022	#####	0000-00-0
30	1	36	6	0	2	1	0	3 70100115	2	3	3	2	1	1	701001	2/1/2022	2/1/2022	#####	0000-00-0
30	2	43	7	15	2	1	0	1 86030510	2	3	3	1	4	4	860305	2/1/2022	2/1/2022	#####	0000-00-0

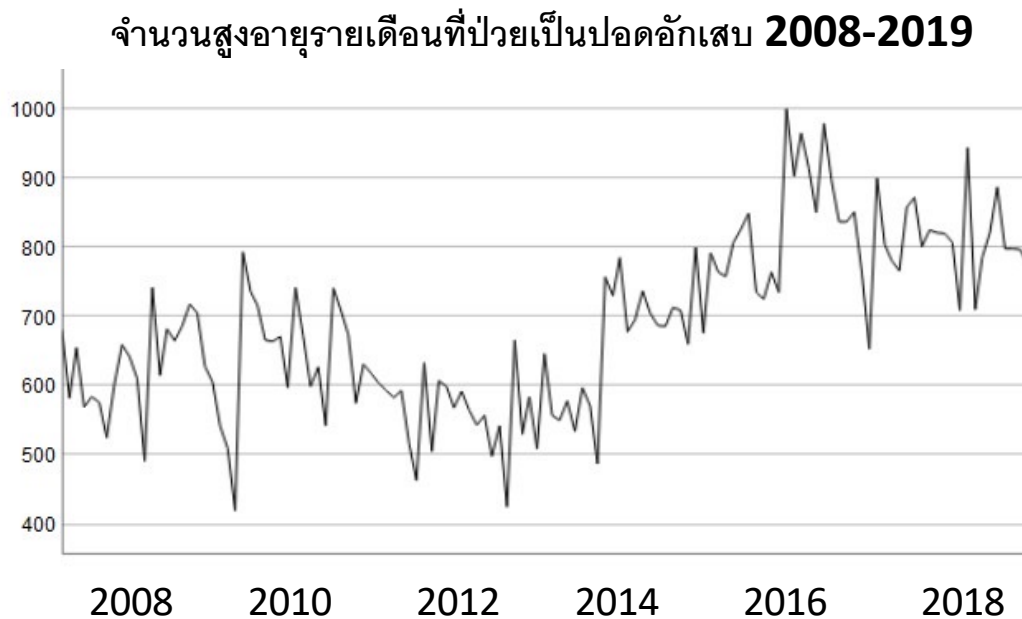
Is this dataset time series data?

disease	sex	agey	agem	aged	marietal	race	race1	occupat	addrcode	metropol	hospital	type	result	hserv	datesick	datedefinida
30	1	24	5	24	1	1	0	3	85030404	2	3	1	3	85030	1/1/2022	1/1/2022 3/
30	1	54	0	0	2	1	0	1	86060209	1	3	2	1	86060	1/1/2022	1/1/2022 ##
30	1	63	5	13	2	1	0	1	85030406	2	3	1	3	85030	1/1/2022	1/1/2022 3/
30	1	11	6	4	1	3	1	6	63080202	2	3	1	1	63080	1/1/2022	2/1/2022 5/
30	1	27	10	10	2	1	0	3	85050105	1	3	1	1	85050	1/1/2022	2/1/2022 4/
30	1	5	1	3	1	1	0	11	63080301	2	3	1	1	63080	1/1/2022	3/1/2022 5/
30	1	4	4	0	1	1	0	11	63050403	2	3	1	3	63050	1/1/2022	3/1/2022 ##
30	1	30	0	0	1	1	0	1	82070601	2	4	1	1	82070	1/1/2022	4/1/2022 5/
30	1	24	0	0	1	1	0	1	82070601	2	4	1	1	82070	1/1/2022	4/1/2022 5/
30	1	25	8	16	1	1	0	1	85050201	1	3	1	1	85050	1/1/2022	5/1/2022 7/
30	1	16	0	8	1	3	1	1	63080409	1	3	1	1	63080	1/1/2022	6/1/2022 9/
30	2	24	0	8	1	3	1	11	63080303	2	3	1	1	63080	1/1/2022	6/1/2022 9/
30	1	28	0	0	2	1	0	10	86010208	1	3	2	1	86010	1/1/2022	6/1/2022 ##
30	1	13	3	23	1	1	0	6	85050201	1	3	1	1	85050	1/1/2022	7/1/2022 9/
30	1	15	1	0	1	1	0	6	63050302	2	3	1	3	63050	1/1/2022	8/1/2022 ##
30	1	15	0	29	1	1	0	6	63080202	2	3	1	1	63080	1/1/2022	9/1/2022 ##
30	2	8	0	0	1	1	0	6	95040307	2	3	1	3	95040	1/1/2022	9/1/2022 ##
30	1	25	5	5	2	1	0	15	85030208	2	3	1	3	85030	1/1/2022	9/1/2022 ##
30	1	36	6	0	2	1	0	3	70100115	2	3	2	1	70100	2/1/2022	2/1/2022 ##
30	2	43	7	15	2	1	0	1	86030510	2	3	1	4	86030	2/1/2022	2/1/2022 ##

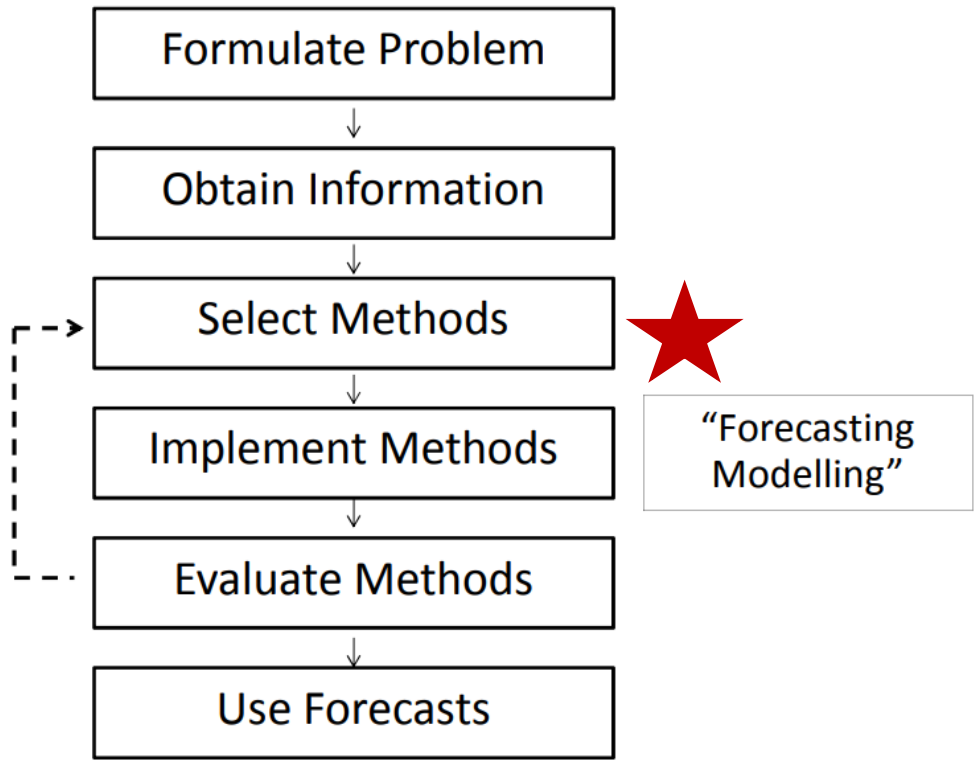
+ ม.ค.	111
+ ก.พ.	56
+ มี.ค.	69
+ เม.ย.	152
+ พ.ค.	340
+ มิ.ย.	508
+ ก.ค.	415
+ ส.ค.	282
+ ก.ย.	89

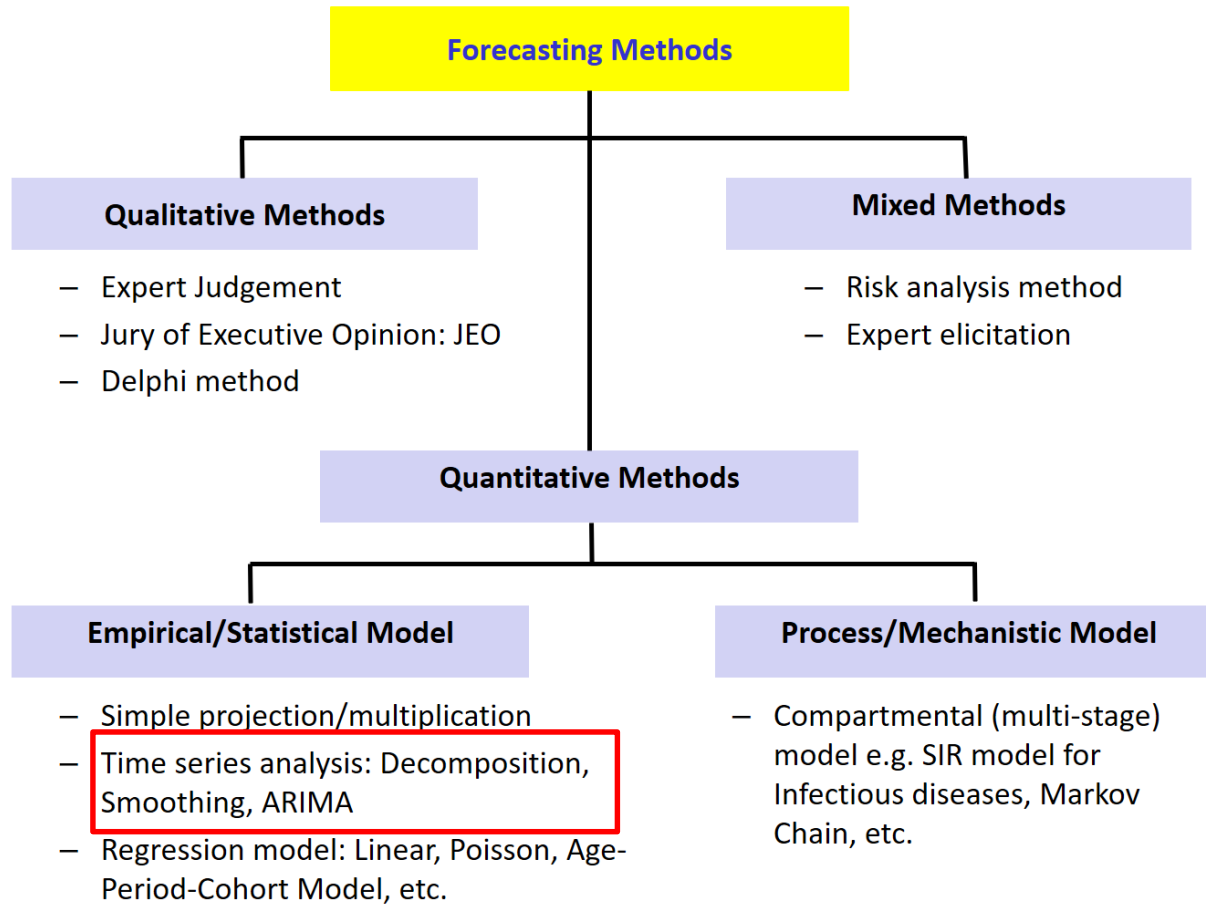
Is this dataset time series data?

⊕ ม.ค.	111
⊕ ก.พ.	56
⊕ มี.ค.	69
⊕ เม.ย.	152
⊕ พ.ค.	340
⊕ มิ.ย.	508
⊕ ก.ค.	415
⊕ ส.ค.	282
⊕ ก.ย.	89

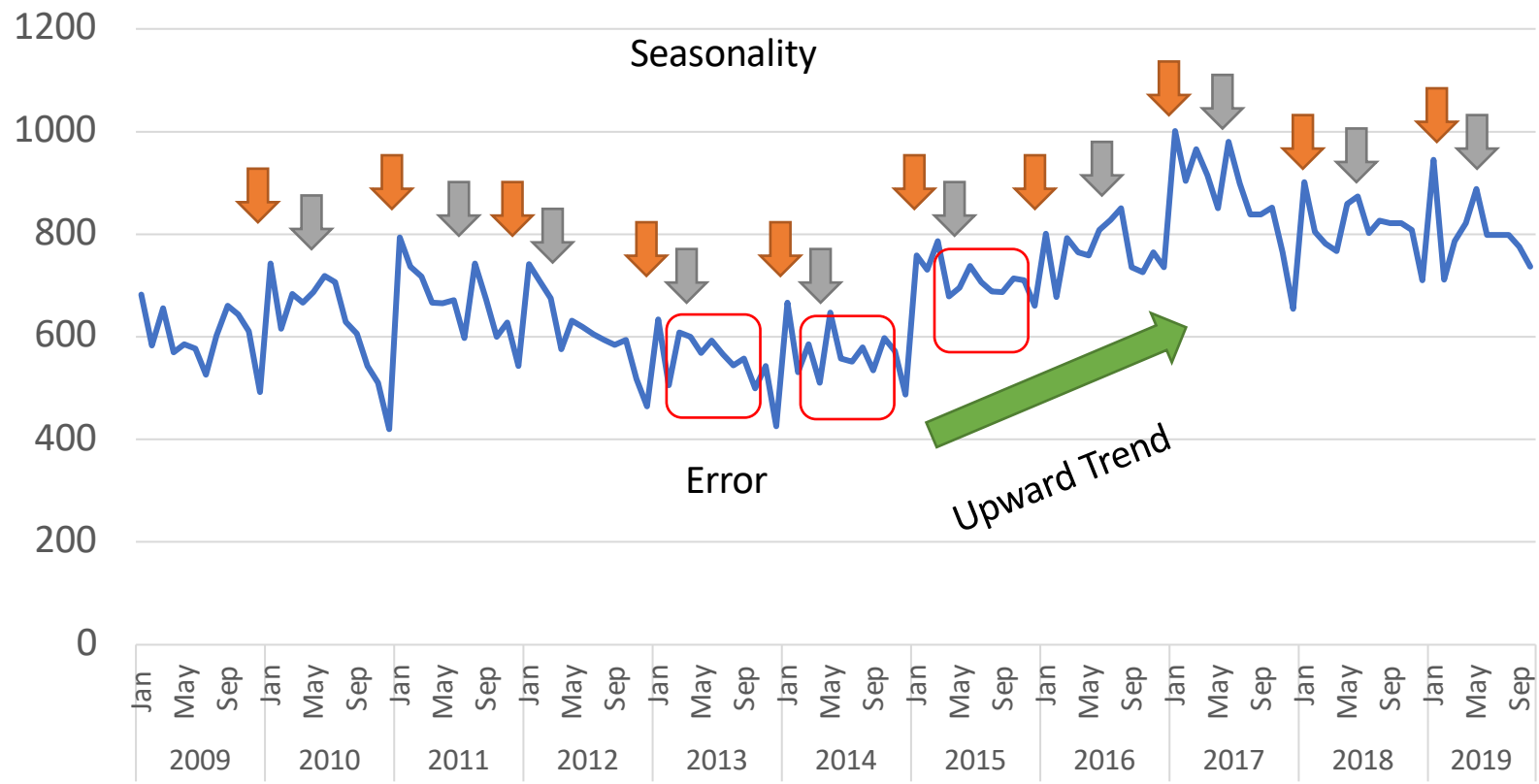


ขั้นตอนการพยากรณ์โรค





จำนวนสูงอายุนรายเดือนที่ป่วยเป็นปอดอักเสบ 2008-2019

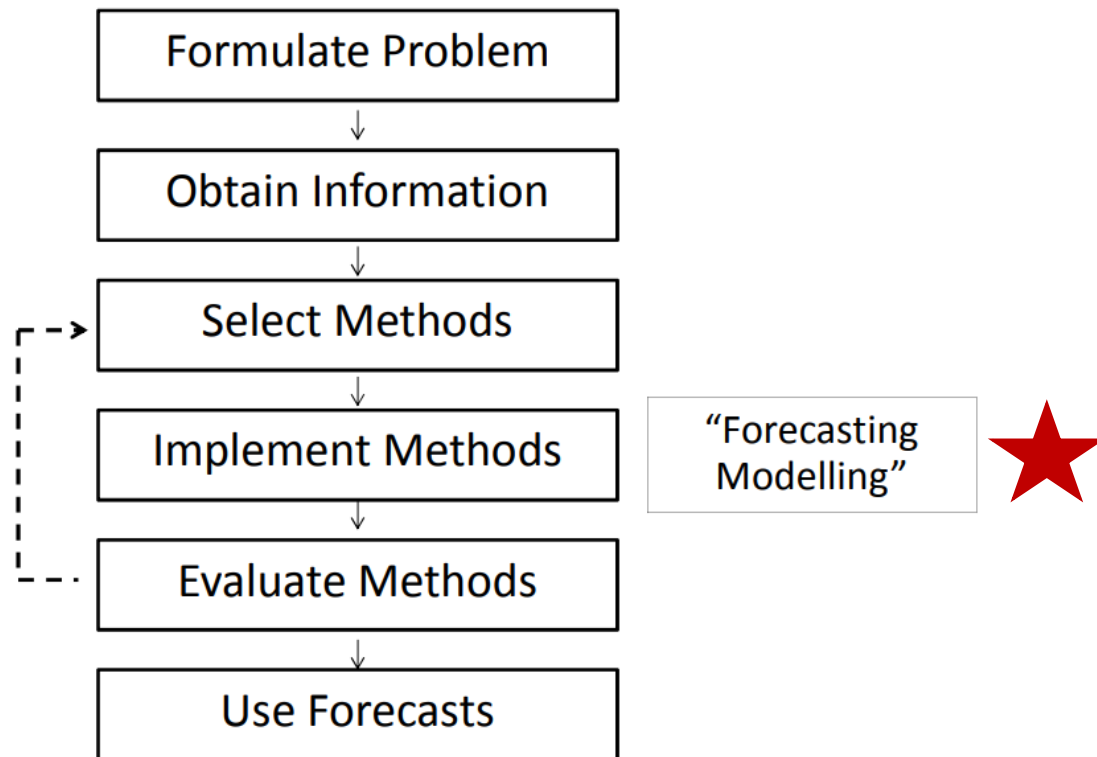


Select Methods

1. Select the forecasting method to use by first **plotting** out the past data to see if there is trend and seasonality. (Plot กราฟ)
2. For each **appropriate method**, determine the **optimized parameters** by minimizing SSE. (เลือกวิธีการ และ **parameters** ที่เหมาะสม)
3. Compare selected methods using quantitative measures. (เปรียบเทียบความแม่นยำของแต่ละวิธีการ)

Method	Smoothing Function	Trend (T)	Seasonality (S)	Holiday Period	Multivariable	Parameters (optimized)
Moving Average	✓					m
Single ES	✓					α
Double ES	✓	✓				α, β
Triple ES (Add)	✓	✓	✓			α, β, γ
Triple ES (Mult)	✓	✓	✓			α, β, γ
ARIMA	✓	✓				p,d,q
SARIMA	✓	✓	✓			p,d,q P,D,Q
Prophet	✓	✓	✓	✓		
VAR	✓	✓	✓		✓	

ขั้นตอนการพยากรณ์โรค

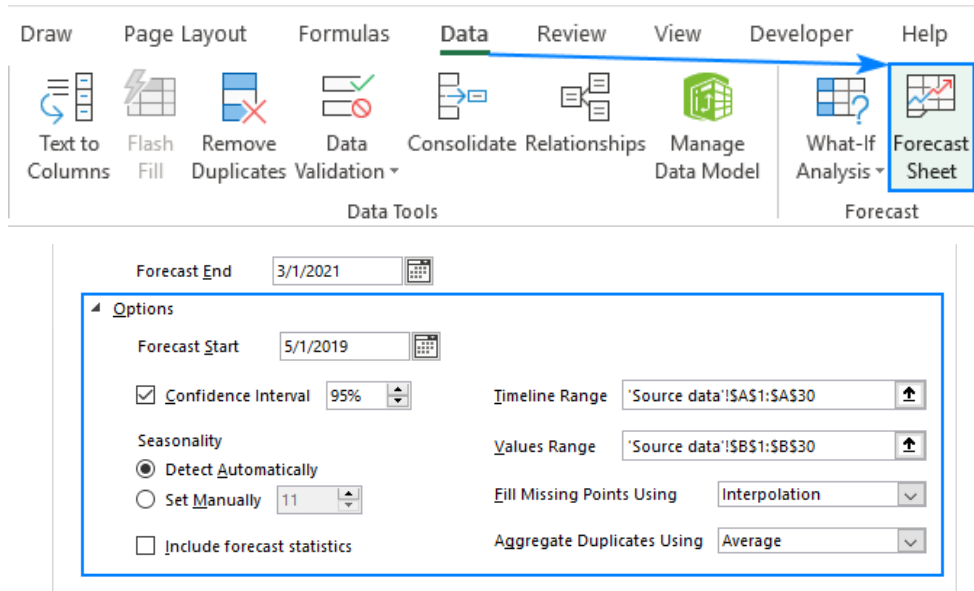


Implement Methods: Software

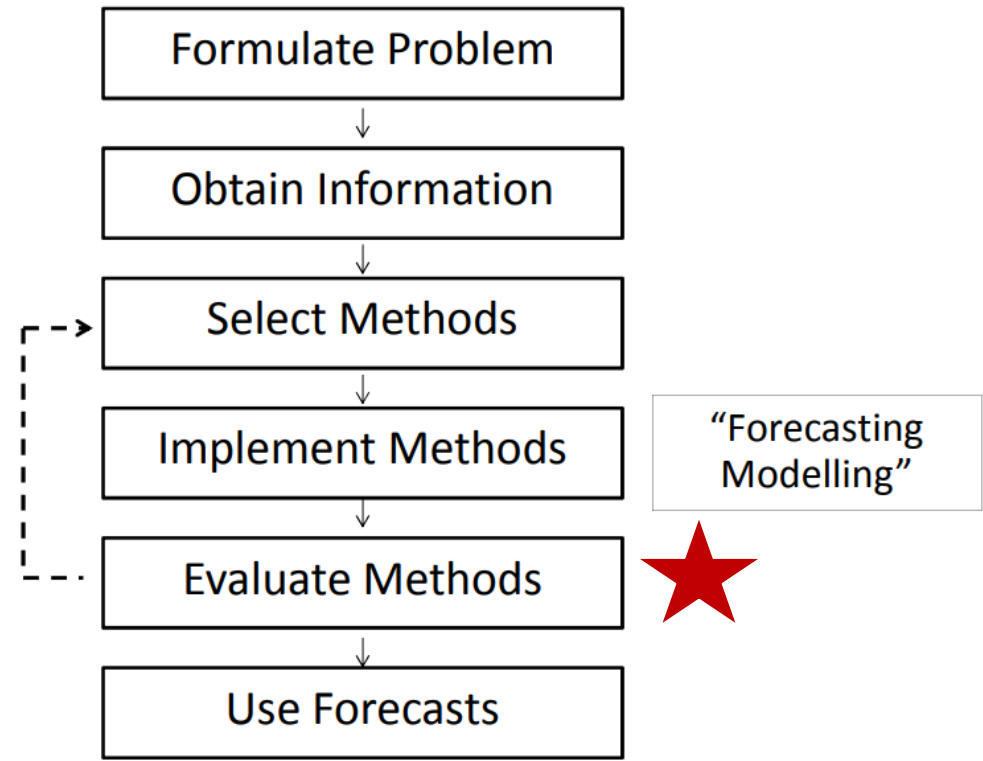


การใช้ Forecast Sheet ใน Excel

- Exponential smoothing forecasting in Excel is based on the *Triple Exponential Smoothing (Additive)* algorithm.

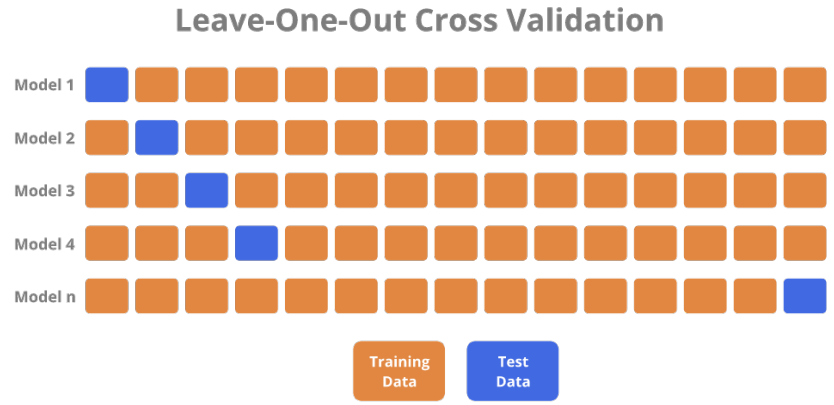
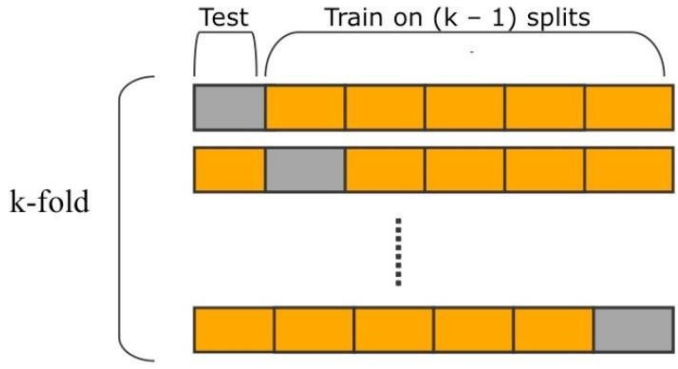
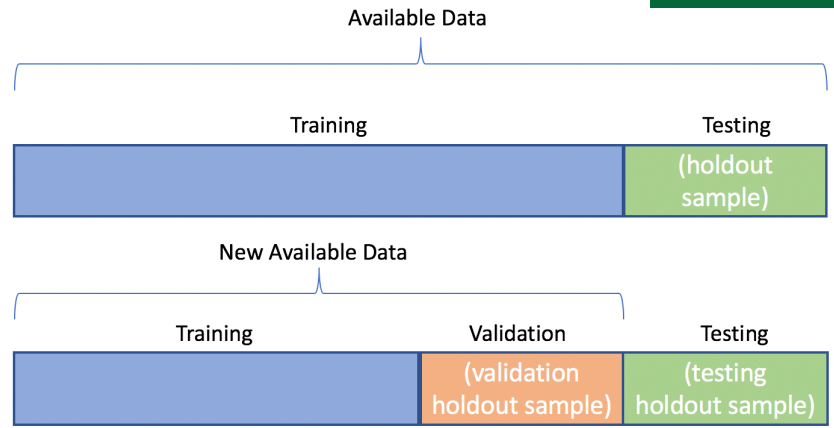


ขั้นตอนการพยากรณ์โรค



Model Evaluation

- Split Train and Test
- K-fold Cross validation
- Leave One Out Cross Validation (LOOCV)



ประเมินประสิทธิภาพ Model (Evaluate Model)

- Minimizing error (ลดความคลาดเคลื่อน)

$$\begin{aligned} \text{ค่าความคลาดเคลื่อน (E}_t\text{)} &= \\ \text{ค่าจริง (A}_t\text{)} - \text{ค่าทำนาย (F}_t\text{)} \end{aligned}$$

Measurements for forecast accuracy (1)

1) Mean Error (ME) $= \frac{\sum (E_t)}{n}$

2) Mean Absolute Error (MAE) $= \frac{\sum |E_t|}{n}$

3) Mean Percent Error (MPE) $= \frac{\sum \left(\frac{E_t}{A_t} \right)}{n}$

4) Mean Absolute Percent Error (MAPE) $= \frac{\sum \left| \frac{E_t}{A_t} \right|}{n}$

Measurements for forecast accuracy (2)

5) Mean Squared Error (MSE) $= \frac{\sum (E_t)^2}{n}$

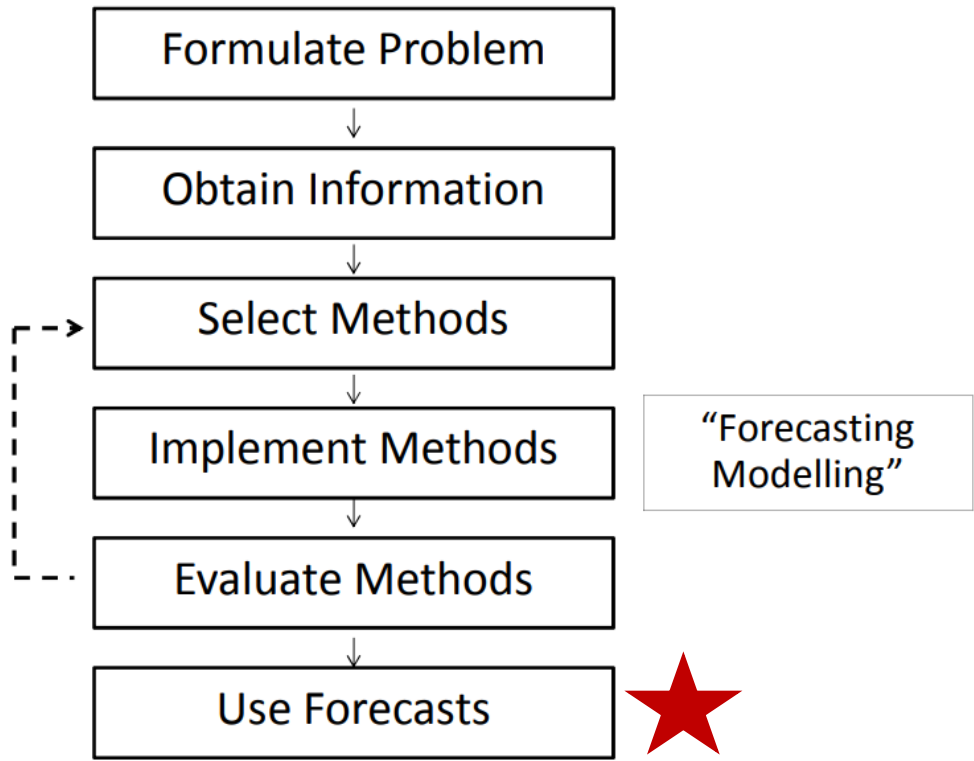
6) Root-Mean Squared Error (RMSE) $= \sqrt{\frac{\sum (E_t)^2}{n}}$

7) Theil's U $= \frac{\sqrt{\sum (A_t - F_t)^2}}{\sqrt{\sum (A_t - A_{t-1})^2}}$

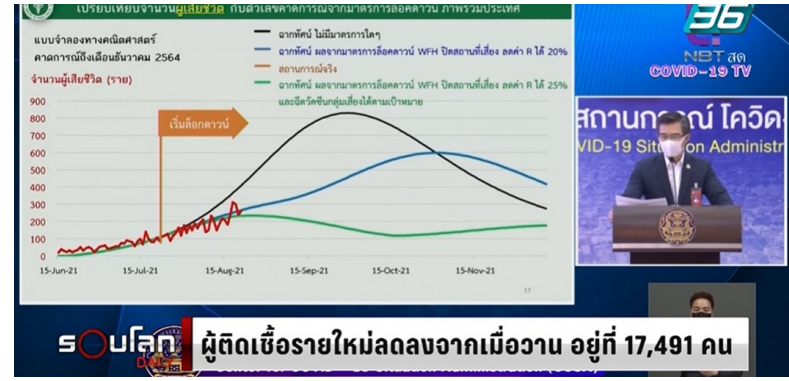
Sometimes called uncertainty coefficient

Note: Other goodness of fit tests such as AIC and BIC are also useful.

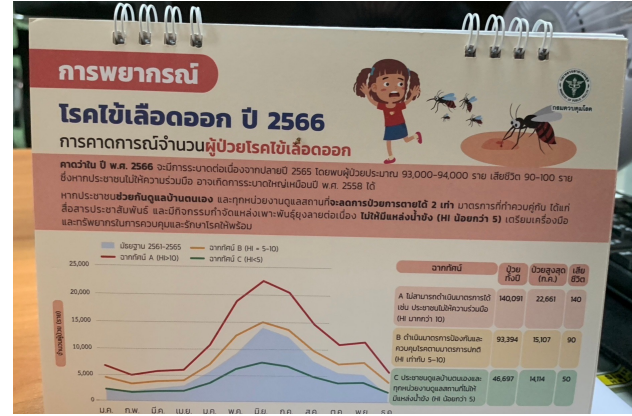
ขั้นตอนการพยากรณ์โรค



กลุ่มเป้าหมาย และ product จากการพยากรณ์ของหน่วยงาน



สำนักงานสาธารณสุขจังหวัด



การพยากรณ์

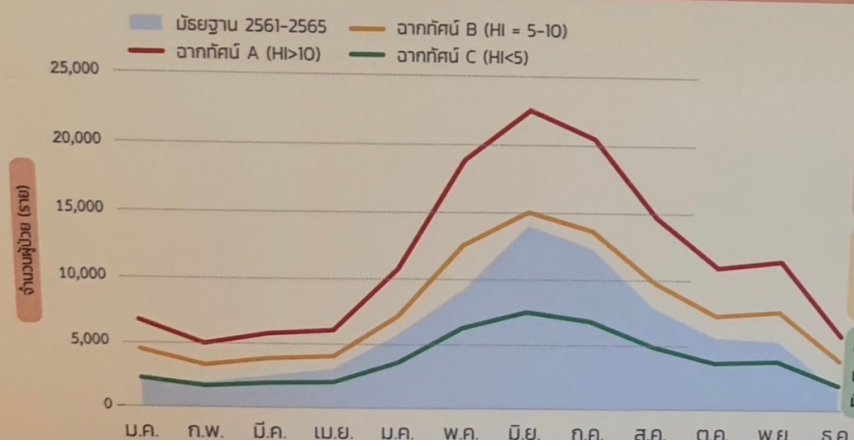
โรคไข้เลือดออก ปี 2566

การคาดการณ์จำนวนผู้ป่วยโรคไข้เลือดออก



คาดว่าใน ปี พ.ศ. 2566 จะมีการระบาดต่อเนื่องจากปลายปี 2565 โดยพบผู้ป่วยประมาณ 93,000-94,000 ราย เสียชีวิต 90-100 ราย ซึ่งหากประชาชนไม่ให้ความร่วมมือ อาจเกิดการระบาดใหญ่เหมือนปี พ.ศ. 2558 ได้

หากประชาชนช่วยกันดูแลบ้านตนเอง และทุกหน่วยงานดูแลสถานที่ที่จะลดการป่วยการตายได้ 2 เท่า มาตรการที่ทำความคุ้นเคย ได้แก่ สื่อสารประชาสัมพันธ์ และมีกิจกรรมกำจัดแหล่งเพาะพันธุ์ยุงลายต่อเนื่อง **ไม่ให้มีแหล่งน้ำขัง (HI น้อยกว่า 5)** เตรียมเครื่องมือและทรัพยากรในการควบคุมและรักษาโรคให้พร้อม




จากทัศน	ป่วยทั้งปี	ป่วยสูงสุด (ก.ค.)	เสียชีวิต
A ไม่สามารถดำเนินมาตรการได้ เช่น ประชาชนไม่ให้ความร่วมมือ (HI มากกว่า 10)	140,091	22,661	140
B ดำเนินมาตรการป้องกันและควบคุมโรคตามมาตรการปกติ (HI เท่ากับ 5-10)	93,394	15,107	90
C ประชาชนดูแลบ้านตนเองและทุกหน่วยงานดูแลสถานที่ไม่ให้มีแหล่งน้ำขัง (HI น้อยกว่า 5)	46,697	14,114	50



Beyond Predictive Modeling: Cost Effectiveness Analysis

Concept of Scarcity



Money is a way of
creating scarcity.

Peter Coyote

quote fancy

Scarcity

[ˈsker-sə-tē]

A basic economics problem—the gap between limited resources and theoretically limitless wants.

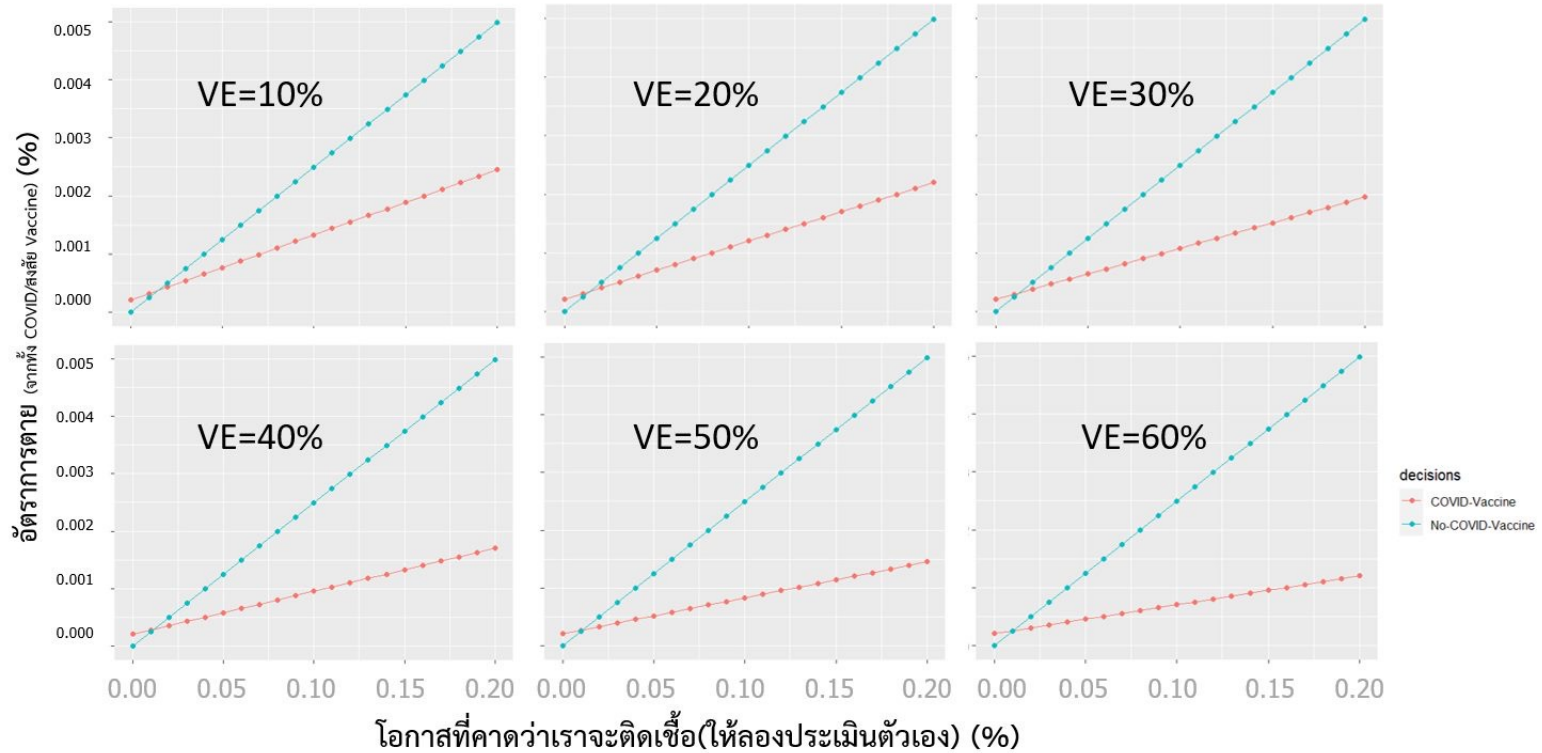


Cost Effectiveness Analysis

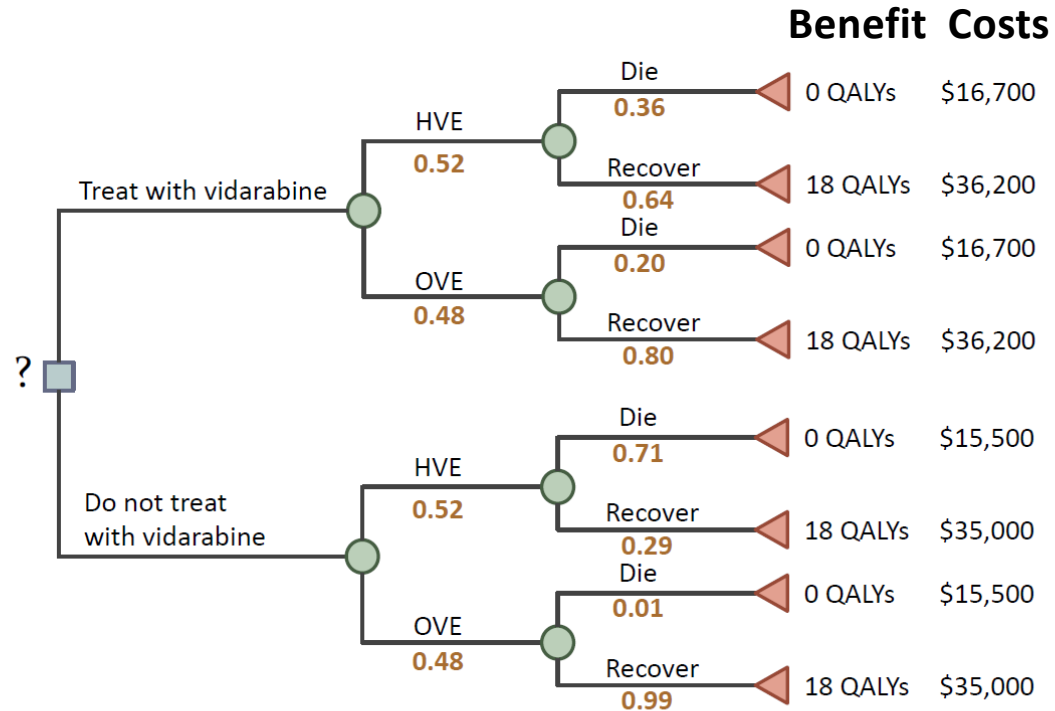
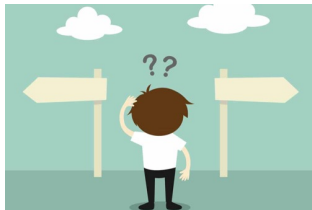
- Used to evaluate the **costs** and **benefits** of **different healthcare interventions**.
- Can help policymakers and healthcare providers make decisions about how to
 - allocate resources
 - prioritize interventions based on their potential impact on health outcomes.

Decision Tree

เปรียบเทียบตามประสิทธิภาพวัคซีน (VE)

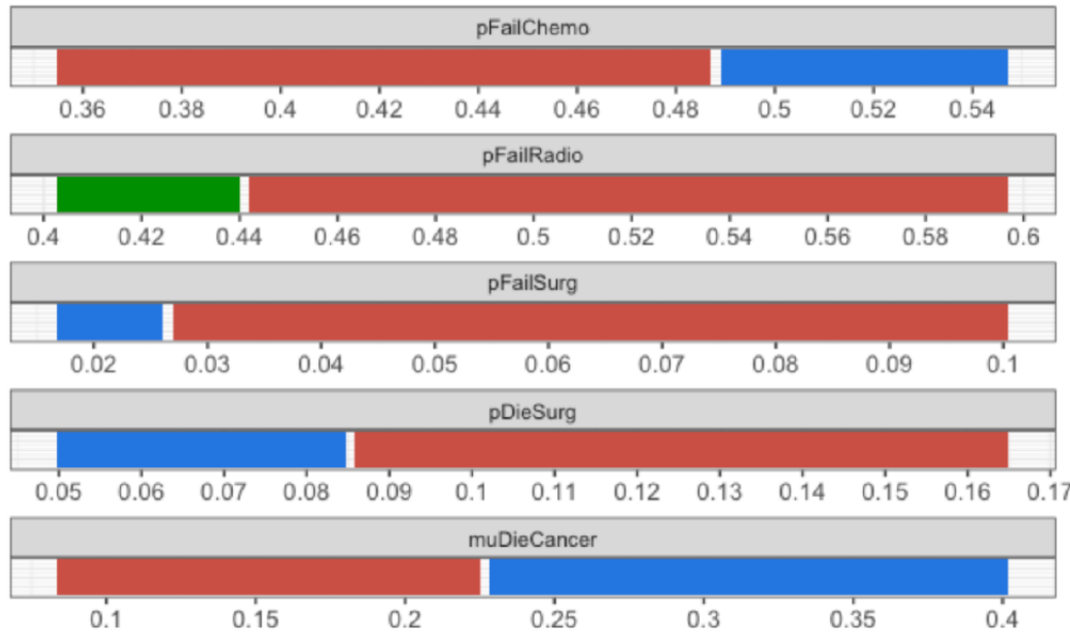


Benefits and Costs



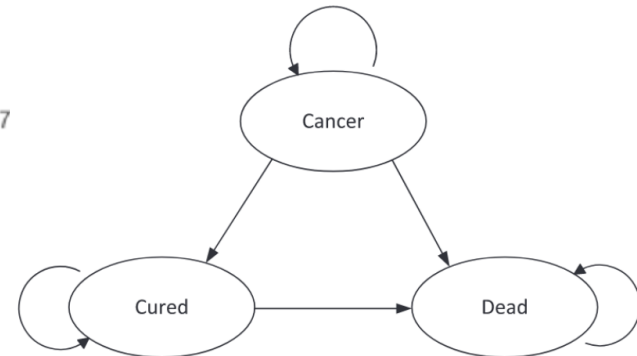
Dealing with uncertainty: One-Way Sensitivity Analysis

- Vary 1 parameter at a time, keeping all others fixed



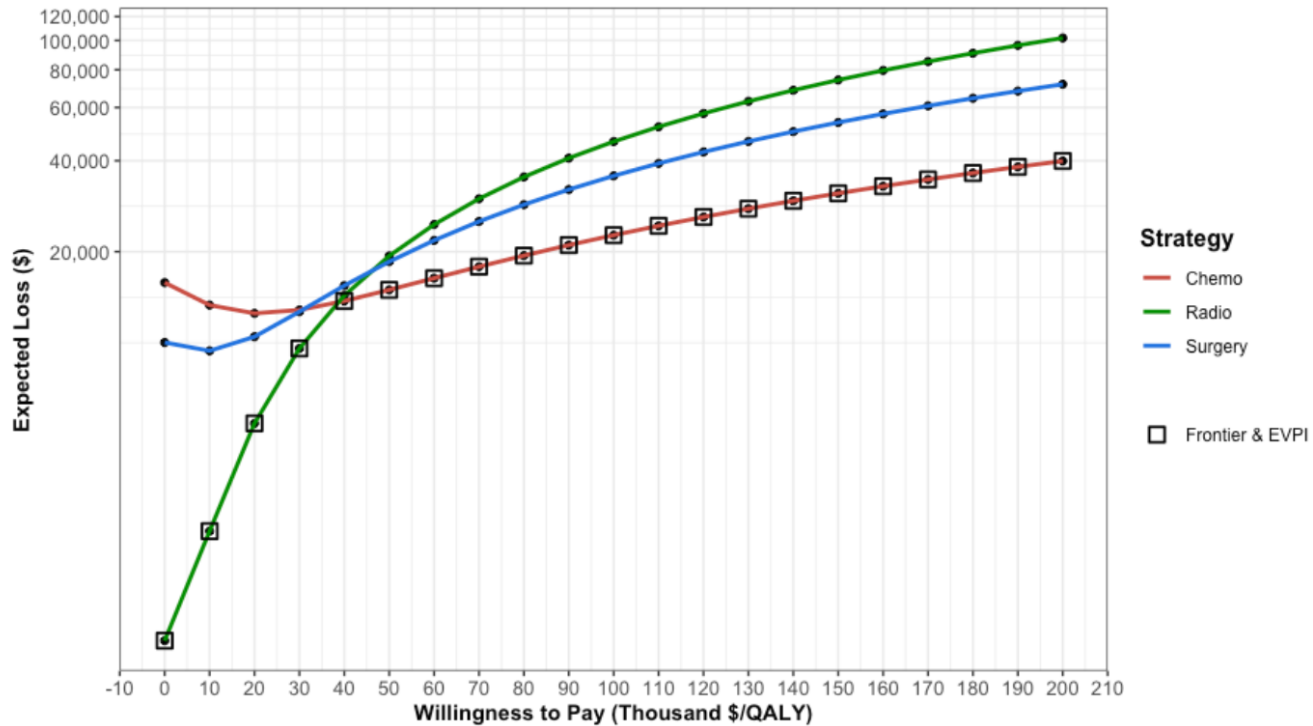
Using WTP = \$100,000 per QALY gained

Optimal Strategy:



Cost Effectiveness Analysis (Frontier)

- Expected loss for each strategy for different WTP



Thank you