ARIMA

For Time-Series Analysis

⇒ **AutoRegressive (AR) model**

idea: Today's value depends on "past value + noise"

autoregressive coefficient

$$\boxed{y_t = c + \phi_1 y_{t-1} + e_t} + \phi_2 y_{t-2} \cdots$$

- y_t: Series at time 't'
- c: intercept (constant mean effect)
- e_t: error at t | $\sim N(0, \sigma^2)$ white noise

Represented by Parameter (P)

NOTE: if $|\phi| < 1$, the series is stationary (fluctuations at constant mean)

Eg: If $\phi = 0.8$,

then y_T is persistent (Today ≈ yesterday)

⇒ Differencing (I) model

Represented by parameter (d) } no. of differences

- Involves transformation a non stationary time series into stationary one by differencing consecutive observations.

- Differencing can be applied multiple time until time series becomes stationary

$$\boxed{x'_t = y_t - y_{t-1}}$$

x'_t → differenced time series at time t

y_t → original series at time t

y_{t-1} → value of series at previous time step.

⇒ Moving average model (MA)

Represented by 'q'

- Indicates dependence of current observation on previous forecast errors

$$\boxed{y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}}$$

y_t → current observation

e_t → error at time 't'

$\underbrace{\theta_1 \rightarrow \theta_q}_{\text{moving avg. parameters}}$

Eg

t	date	y_t
1	2025-1-1	100
2	2025-1-2	102
3	2025-1-3	101
4	2025-1-4	103
5	2025-1-5	106
6	2025-1-6	108

ARIMA(1,1,0)

You're given a short time series y_t

Sol^n

Step 1: compute differences

$$x_t = y_t - y_{t-1} \quad (\text{for } t=2 \text{ to } t=6)$$

$$x_2 = y_2 - y_1 = 102 - 100 = 2$$

$$x_3 = y_3 - y_2 = 101 - 102 = -1$$

$$x_4 = y_4 - y_3 = 103 - 101 = 2$$

$$x_5 = y_5 - y_4 = 106 - 103 = 3$$

$$x_6 = y_6 - y_5 = 108 - 106 = 2$$

make ordered pairs (x_{t-1}, x_t) for $t=3$ to $t=6$

$(2,-1)\ (-1,2)\ (2,3)\ (3,2)$

Step 2 : Calculate ...

(i) sample means of ordered pairs $(\bar{x})$ & $(\bar{y})$

(ii) Covariance & variance S_{xy} & S_{xx}

(iii) Slope $\alpha = \frac{S_{xy}}{S_{xx}}$

(iv) intercept $c = \bar{y} - \alpha\bar{x}$

Pairs $\rightarrow$ $x_1 = 2$ $y_1 = -1$

$x_2 = -1$ $y_2 = 2$

$x_3 = 2$ $y_3 = 3$

$x_4 = 3$ $y_4 = 2$

(i) $\bar{x} = \frac{6}{4} = 1.5$ // $\bar{y} = \frac{6}{4} = 1.5$ //

(ii) $S_{xy} = \sum_{i=1}^{m} (x_i - \bar{x})(y - \bar{y}_i)$ $S_{xx} = \sum_{i=1}^{m} (x_i - \bar{x})^2$

$S_{xy} = [(0.5)(-2.5) + (-2.5)(0.5) + (0.5)(1.5) + (1.5)(0.5)]$

$S_{xx} = [(0.5)^2 + (-2.5)^2 + (0.5)^2 + (1.5)^2]$

$S_{xx} = 9$ //

$S_{xy} = -1$ //

(iii) $\alpha = \frac{S_{xy}}{S_{xx}} = \frac{-1}{9} = -0.111\ldots$ //

(iv) $c = \bar{y} - \alpha\bar{x} = 1.5 - \left(-\frac{1}{9}\right)(1.5) = 1.666\ldots$ //

$\downarrow$

$5/3$

Step 4 : Find (i) Fitted values

$$\hat{x}_t = c + \phi x_{t-1}$$

& (ii) Residuals

$$e_t = x_t - \hat{x}_t$$

$t=3$

$$\hat{x}_3 = \frac{5}{3} + \left(\frac{-1}{9}\right)2 = \frac{13}{9}$$

$t=4$

$$\hat{x}_4 = \frac{5}{3} + \left(\frac{-1}{9}\right)(-1) = \frac{16}{19}$$

$t=5$

$$\hat{x}_5 = \frac{5}{3} + \left(\frac{-1}{9}\right)3 = \frac{13}{9}$$

$t=6$

$$\hat{x}_6 = \frac{5}{3} + \left(\frac{-1}{9}\right)2 = \frac{4}{3}$$

$t=3$

$$e_3 = -1 - \frac{13}{9} = \frac{-22}{9}$$

$t=4$

$$e_4 = 2 - \frac{16}{9} = \frac{2}{9}$$

$t=5$

$$e_5 = 3 - \frac{13}{9} = \frac{14}{9}$$

$t=6$

$$e_6 = 2 - \frac{4}{3} = \frac{2}{3}$$

Step 5 : SSR & estimate of σ^2

$$e_3^2 = 5.97$$

$$e_4^2 = 0.049$$

$$e_5^2 = 2.419$$

$$e_6^2 = 0.444$$

$$SSR = e_3^2 + e_4^2 + e_5^2 + e_6^2 = 8.89$$

$$DoF = m - K \xrightarrow{} \text{estimated parameters}$$
$$= 4 - 2$$
$$= 2$$

$$\sigma^2 = \frac{SSR}{DoF} = \frac{8.89}{2} = 4.44$$

$$\sigma = 2.108$$

Step 6: one-step Forecast (level)

one step forecast for next difference

$\rightarrow \hat{x}_{7,6} = \phi x_6 + \hat{c}$

$x_{7,6} = \left(\frac{-1}{9}\right)(2) + \frac{5}{3}$

$x_{7,6} = \frac{13}{9}$

The level Forecast:

$Y_{7,6} = Y_6 + x_{7,6}$

$Y_{7,6} = 108 + \frac{13}{9} = \boxed{109.444..}$ //

NOTE: The calculations changes as p,d,q value changes

$p \rightarrow$ How many past values we use

$d \rightarrow$ How many times we difference the data

$q \rightarrow$ how many past errors we use

General model:

$$\boxed{(1-B)^d y_t \cdot \phi(B) = e_t \cdot \theta(B)}$$

$B \rightarrow$ Backshift operator
$(BY_T = Y_{t-1})$

e_t = error

$\phi(B) \rightarrow 1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p$

$\theta(B) \rightarrow 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^p$

Eg $y = [5, 7, 10, 15, 22, 32]$

ARIMA(0,2,0) predict y_7 & y_8

Soln

$p=0$; no past values are used

$q=0$; no past errors are used $\longrightarrow e_t = 0$

$d = 2$

$$y_{t+1} - y_t - y_t + y_{t-1} = 0$$

$$y_{t+1} - 2y_t + y_{t-1} = 0$$

$$y_{t+1} = 2y_t - y_{t-1}$$

$y_7 = 2y_5 - y_4$
$= 2(32) - 15$
$= 42$

$y_8 = 2(y_7) - y_6$
$= 2(42) - 32$
$= 52$