

Demo phân tích chuỗi thời gian

Ngày 9 tháng 6 năm 2018

- 1 Mô tả dữ liệu
- 2 Xử lý chuỗi thời gian
- 3 Ước lượng khoảng cách chuỗi thời gian

- Là dữ liệu mưa theo tháng tại trạm Tam Kỳ tỉnh Quảng Nam.
- Thời gian theo dõi từ 2008-2016: chia thành 2 giai đoạn 2008-2012 và 2013-2016.

- Đọc dữ liệu

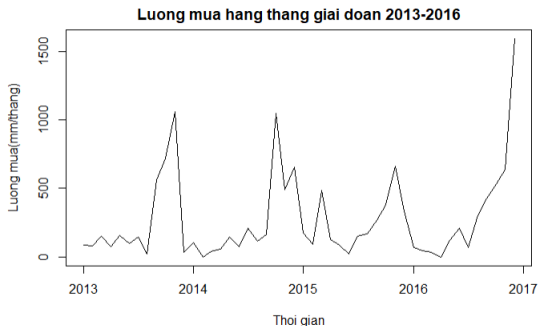
```
data_frame <- read.csv("Mua_thang_TK.csv")  
head(data_frame)
```

	Nam <int>	Thang <int>	TK <dbl>
1	1977	1	0.0
2	1977	2	50.0
3	1977	3	51.0
4	1977	4	0.0
5	1977	5	61.1
6	1977	7	85.8

Dữ liệu mưa tháng giai đoạn 2013-2016

- Dữ liệu lượng mưa giai đoạn 2013-2016

```
>data_frame_2 <-data_frame[data_frame[, "Nam"] >= 2013  
                           & data_frame[, "Nam"] <= 2016, "TK"]  
>time_series_2 <- ts(data_frame_2[, "TK"], frequency = 12)  
>plot(time_series_2)
```



Phân tích dữ liệu mưa giai đoạn 2013-2016

- Kiểm tra tính dừng

```
>adf.test(time_series_2)
```

Augmented Dickey-Fuller Test

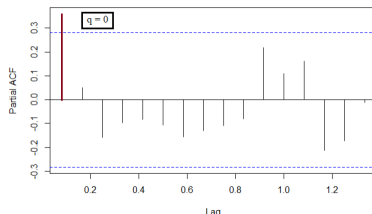
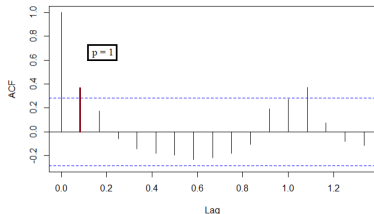
```
data: time_series_2  
Dickey-Fuller = -2.1797, Lag order = 3, p-value = 0.5024  
alternative hypothesis: stationary
```

$d = 0$

- Hệ số ACF và PACF

```
>acf(time_series_2)
```

```
>pacf(time_series_2)
```



Ước lượng tham số cho mô hình

- Các hệ số (p, d, q) lần lượt là $(1, 0, 0)$.
- Mô hình chuỗi thời gian cần ước lượng là AR(1).

```
>fit2 <- Arima(time_series_2, order = c(1, 0, 0))  
>fit2
```

```
Series: time_series_2  
ARIMA(1,0,0) with non-zero mean
```

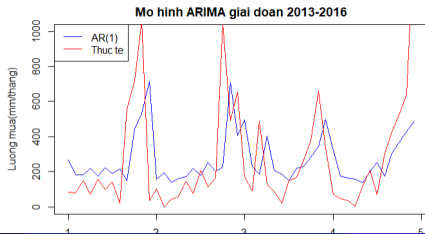
```
Coefficients:
```

```
      ar1      mean  
    0.5459 303.2004  
s.e.  0.1558  90.4153
```

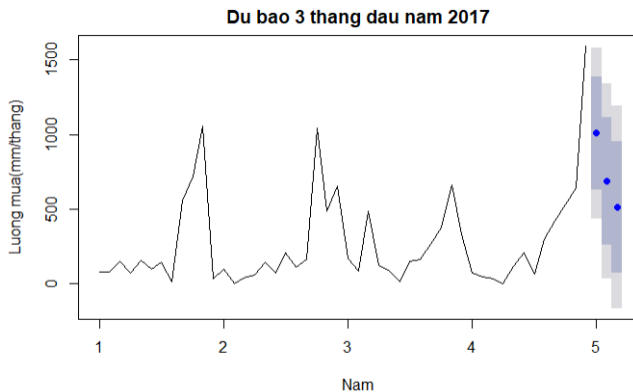
```
sigma^2 estimated as 85844: log likelihood=-339.91  
AIC=685.82  AICc=686.37  BIC=691.44
```

```
Training set error measures:
```

```
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1  
Training set 3.236544 286.8227 196.9081 -Inf  Inf  0.9804959 -0.01254458
```



```
>forecast <- forecast(fit2, h = 3)  
>plot(forecast)
```



So sánh mô hình chuỗi thời gian của các giai đoạn 2008-2012, 2013-2016 và 2008-2016

- Chuỗi thời gian `time_series_1` và `time_series` lần lượt tương ứng với giai đoạn 2008-2012 và 2008-2016.
- Hàm `auto.arima` cho phép hỗ trợ phân tích chuỗi thời gian một cách nhanh chóng.

```
>auto.arima(time_series_1)
>auto.arima(time_series_2)
>auto.arima(time_series)
```

	p	d	q	P	D	Q
2008-2012	1	0	0	1	0	0
2013-2016	1	0	0			
2008-2016	0	0	0	2	0	0

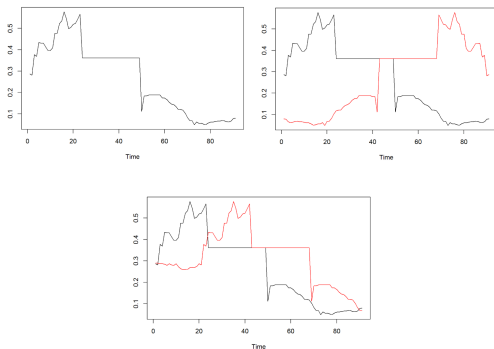
Một số loại khoảng cách giữa chuỗi thời gian

- Model-free:
 - Là các khoảng cách Euclidean, Mahatan.
 - Công thức tổng quát: $d_{L_p}(X_T, Y_T) = (\sum_{t=1}^T (x_t - y_t)^q)^{\frac{1}{q}}$
- Khoảng cách dựa trên hình dạng:
 - Khoảng cách DTW(Dynamic time warping), CCD(Cross-correlation distance)
- Khoảng cách dựa trên thuộc tính:
 - Là các khoảng cách dựa trên hệ số tự tương quan của hai chuỗi thời gian.
 - Khoảng cách PACF, ACF.

Khảo sát một số loại khoảng cách chuỗi thời gian

- Các loại khoảng cách lần lượt được khảo sát theo các trường hợp sau:
 - Hai chuỗi thời gian giống hệt nhau
 - Nghịch đảo
 - Tịnh tiến
- Chuỗi thời gian được chuẩn hóa về đoạn 0-1.
- Khoảng cách càng nhỏ nghĩa là hai chuỗi thời gian đó càng "giống nhau".

Kết quả khảo sát



	EUCL	DTW	CCD	ACF	PACF
TS - TS	0	0	0	0	0
TS - TS Nghịch đảo	2.99	16.98	0.14	0	0
TS - TS Tịnh tiến	1.47	0.3	0.23	1.45	0.56