Tujuan Pembelajaran

Setelah memperlajari modul ini, mahasiswa diharapkan:

1. Mampu memahami dan menjelaskan konsep dasar time series
2. Mampu memodelkan data deret waktu dengan menggunakan metode ARIMA Box-Jenkins non musiman menggunakan software R.
3. Mampu melakukan peramalan (forecasting) untuk beberapa tahap ke depan dengan menggunakan software R.
4. Mampu menginterpretasikan hasil pengolahan data yang diperoleh.

A. Pendahuluan

Peramalan (*forecasting*) dengan menggunakan data deret waktu (*time series*) adalah hal yang penting bagi seseorang yang menghadapi situasi yang tidak menentu. Oleh karena itu, pembahasan terkait pengertian deret waktu dan Model ARIMA yang dikembangkan oleh Box dan Jenkins adalah penting.

B. Materi dan Prosedur

1. Pengertian Deret Waktu

Data deret waktu merupakan serangkaian data pengamatan yang terjadi berdasarkan indeks waktu secara berurutan dengan interval waktu tetap. Analisis deret waktu adalah salah satu prosedur statistika yang diterapkan untuk meramalkan struktur probabilistik keadaan yang akan terjadi di masa yang akan datang dalam rangka pengambilan keputusan.

Suatu urutan pengamatan memiliki model deret waktu jika memenuhi dua hal. Pertama, nterval waktu antar indeks waktu *t* dapat dinyatakan dalam satuan waktu yang sama (identik). Kedua, adanya ketergantungan antara pengamatan Z_t dengan Z_{t+k} yang dipisahkan oleh jarak waktu berupa kelipatan Δ_t sebanyak *k* kali (dinyatakan sebagai lag *k*).

Tujuan analisis deret waktu antara lain adalah (1). meramalkan kondisi di masa yang akan datang (*forecasting*); mengetahui hubungan antarpeubah; dan kepentingan kontrol (untuk mengetahui apakah proses terkendali atau tidak).

2. Model ARIMA Box-Jenkins

Model AutoRegressive Integrated Moving Average (ARIMA) telah dipelajari secara mendalam oleh George Box dan Gwilym Jenkins (1976). Box dan Jenkins secara efektif telah berhasil mencapai kesepakatan mengenai informasi relevan yang diperlukan untuk memahami dan menggunakan model-model ARIMA untuk deret waktu satu peubah (*univariate*). Model ARIMA terdiri dari tiga aspek, yaitu aspek *autoregressive*, *integrated*, dan *moving average*. Secara umum, model ARIMA ini dituliskan dengan notasi ARIMA (p, d, q), dimana p adalah orde dari proses *autoregressive* (AR), d menyatakan pembedaan (*differencing*), dan q menyatakan orde dari proses *moving average* (MA).

3. Metodologi Box-Jenkins

Metodologi Box-Jenkins terdiri dari empat tahap yaitu: identifikasi, penaksiran (*estimation*) parameter, pemeriksaan diagnostik (*diagnostic checking*) dan peramalan (*forecasting*).

a. Tahap Identifikasi

Pada tahap identifikasi ini, langkah awal yang penting untuk dilakukan adalah menentukan apakah data deret waktu sudah stasioner atau belum, baik dalam rata-rata maupun dalam variansi; apakah data mengandung unsur musiman atau tidak. Konsep stasioner secara sederhana dapat diartikan suatu kondisi dimana nilai suatu data tidak jauh berbeda atau mungkin sama dengan data yang lainnya. Deret data dikatakan stasioner pada rata-rata, jika tidak ada perubahan rata-rata yang jelas dari waktu ke waktu pada diagram deret waktu. Selanjutnya, deret data dikatakan stasioner pada variansi, jika tidak ada perubahan variansi (ragam) yang jelas dari waktu ke waktu pada diagram deret waktu. Apabila rata-rata mengalami perubahan dari waktu ke waktu dengan kata lain menyimpang dengan beberapa pola siklus kecenderungan (*trend-cycle*), maka deret waktu tersebut mempunyai rata-rata yang tidak stasioner.

Dalam praktiknya, proses pengujian kestasioneran dalam rata-rata dapat dilakukan dengan efisien menggunakan *Augmented Dickey-Fuller Test* melalui bantuan paket komputer misalnya menggunakan software R. Bentuk visual plot time serie yang disediakan oleh paket komputer dapat juga digunakan untuk memperlihatkan kestasioneran suatu data.

Apabila data belum stasioner dalam rata-rata, diperlukan cara atau metode untuk meng-menstasionerkan data sebelum melangkah lebih lanjut pada pembentukan model. Salah satu cara dapat dicapai melalui proses *differencing*. Bila kondisi stasioner dalam variansi tidak diperoleh, Box & Cox (1964) memperkenalkan transformasi pangkat, $Z_t^{(\lambda)} = \frac{Z_t^{(\lambda)} - 1}{\lambda}$, dimana λ disebut sebagai parameter transformasi.

Apabila kondisi stasioner baik dalam rata-rata maupun dalam variansi sudah dipenuhi, langkah selanjutnya adalah membuat plot fungsi autokorelasi (*Autocorrelation Function* (ACF)). Plot ACF juga dapat digunakan sebagai alat untuk mengidentifikasi kestasioneran data. Jika nilai-nilai autokorelasi turun secara drastis sampai nol setelah *time-lag* dua atau tiga, ini salah satu indikasi bahwa data sudah stasioner dalam rata-rata. Sebaliknya, jika diagram ACF cenderung turun lambat atau turun secara linear, maka dapat disimpulkan data belum stasioner dalam rata-rata.

Langkah selanjutnya dalam tahap identifikasi adalah membuat plot fungsi autokorelasi parsial (*Partial Autocorrelation Function* (PACF)). Tujuan perhitungan autokorelasi parsial adalah membantu menetapkan model ARIMA yang paling tepat untuk peramalan. Melalui perhitungan ACF dan PACF, dapat ditentukan orde dari *autoregressive* (AR), *moving average* (MA) Jadi, melalui tahap identifikasi, kita dapat menentukan model dugaan awal ARIMA (p, d, q).

Identifikasi model tidak hanya dilakukan dengan melihat plot time series, plot ACF dan PACF, tetapi harus pula disertai dengan pengetahuan mengenai data yang akan dianalisis. Berdasarkan plot time series, plot ACF dan PACF, serta pengetahuan yang cukup mengenai data, model yang akan dibuat dapat menggunakan parameter sesedikit mungkin (prinsip parsimoni).

b. Penaksiran Parameter

Langkah selanjutnya setelah diperoleh dugaan model awal ARIMA (p, d, q) adalah menaksir parameter dari model. Beberapa metode dalam penaksiran parameter model ARIMA Box-Jenkins yaitu metode *moment*, metode *least squared*, metode *maximum likelihood*, dan sebagainya.

c. Pemeriksaan Diagnostik

Pemeriksaan diagnostik (*diagnostic checking*) terbagi menjadi dua bagian yaitu uji kesignifikanan parameter dan uji kesesuaian model. Pengujian kesignifikanan parameter dengan uji *t*, sedangkan uji kesesuaian model meliputi uji asumsi residual white noise dan berdistribusi normal. Pengujian apakah residual *white noise* atau tidak dapat digunakan uji *Ljung-Box*, sedangkan pengujian apakah residual berdistribusi normal atau tidak, digunakan uji *Kolmogorov-Smirnov*

d. Peramalan

Tahap peramalan dapat dilakukan jika seluruh parameter model signifikan dan seluruh asumsi residual terpenuhi.

4. MODEL ARIMA NON-MUSIMAN

Secara umum model ARIMA Box-Jenkins atau ARIMA (p, d, q) adalah sebagai berikut:

$$\emptyset_p(B)(1-B)^d Z_t = \theta_q(B)a_t$$

Dimana

$\emptyset_p(B) = (1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_P B^P)$

$\theta_q(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right);$

$\emptyset_1, \emptyset_2, \emptyset_3, \ldots, \emptyset_p$ adalah koefisien orde p;

$\theta_1, \theta_2, \theta_3, \ldots \theta_q$ adalah koefisien orde q;

$(1-B)^d$ adalah orde *differencing* non-musiman.

p adalah orde AR

d adalah orde *differencing* non-musiman

q adalah orde MA

Z_t adalah besarnya pengamatan (kejadian) pada waktu ke-*t*

a_t adalah suatu proses *white noise* atau galat pada waktu ke-*t* yang diasumsikan mempunyai rata-rata 0 dan variansi konstan σ_a^2.

4.1. *Autoregressive* Orde 1, AR(1) atau ARIMA (1,0,0)

Bentuk umum suatu proses *autoregressive* orde p dinyatakan AR(p) adalah:

$$\emptyset_p(B)Z_t = a_t$$

dimana $\emptyset_p(B) = (1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_P B^P)$

Sehingga model *autoregressive* orde 1, dapat ditulis:

$$(1-\emptyset_1 B)Z_t = a_t$$

Model AR(1) menandakan bahwa orde dari p = 1, d = 0 dan q = 0.

Secara umum rumus ACF untuk AR(1) adalah:

$$\rho_k = \begin{cases} 1 & ; k = 0 \\ \phi_1^{\,k} & ; k > 0 \end{cases}$$

Dari rumus umum ACF untuk model AR(1) di atas terlihat bahwa nilai autokorelasi (ACF) semakin kecil atau mendekati nol seiring bertambahnya *lag* (k). Dapat dikatakan bahwa bentuk ACF dari model AR(1) turun secara eksponensial.

Selanjutnya, secara umum, model AR(1) memiliki bentuk fungsi autokorelasi parsial (PACF) sebagai berikut:

$$\phi_{kk} = \begin{cases} \phi_1; & k = 1 \\ 0; & k > 1 \end{cases}$$

Dari rumus umum PACF untuk model AR(1) di atas terlihat bahwa nilai parsial autokorelasi yang signifikan berbeda dari nol hanya pada *lag 1* saja.

Berdasarkan analisis ACF dan PACF, dapat diperoleh karakteristik dari model AR(1) sebagai berikut.

1. Nilai autokorelasi turun secara eksponensial.
2. Autokorelasi parsial signifikan berbeda dengan nol hanya pada *lag* 1

4.2.*Autoregressive* Orde 2, AR(2) atau ARIMA (2,0,0)

Model *autoregressive* orde 2, dapat ditulis:

$$(1-\emptyset_1 B-\emptyset_2 B^2)Z_t = a_t$$

Model AR(2) menandakan bahwa orde dari p = 2, d = 0 dan q = 0.

Karakteristik dari model AR(2) sebagai berikut.

1. Nilai autokorelasi turun secara eksponensial.
2. Autokorelasi parsial signifikan berbeda dengan nol hanya pada *lag* 1 dan lag 2 (cut off after lag 2).

4.3. *Moving average* orde 1, MA(1) atau ARIMA (0,0,1)

Bentuk umum suatu proses *moving average* orde q dinyatakan MA(q) adalah:

$$Z_t = \theta_q(B)a_t$$

$$\theta_q(B) = (1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)$$

Sehingga model *Moving Average* orde 1, dapat ditulis:

$$Z_t = (1 - \theta_1 B)\, a_t$$

Model MA(1) menandakan bahwa orde dari $p = 0$, $d = 0$ dan $q = 1$.

Secara umum rumus ACF untuk MA(1) adalah:

$$\text{Jadi, } \rho_k = \begin{cases} 1 & ; k = 0 \\ \dfrac{-\theta_1}{1+\theta_1^{\,2}} & ; k = 1 \\ 0 & ; k > 1. \end{cases}$$

Dari rumus umum ACF untuk model MA(1) di atas terlihat bahwa nilai autokorelasi yang signifikan berbeda dari nol hanya pada *lag* 1 saja, atau dapat dikatakan bahwa nilai autokorelasi terpotong sesudah lag 1 (*cut off after lag 1*).

Selanjutnya, secara umum, model MA(1) memiliki bentuk fungsi autokorelasi parsial (PACF) sebagai berikut:

$\emptyset_{kk} = \frac{\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}$, untuk $k = 1, 2, 3, \ldots$,

Dari rumus umum PACF untuk model MA(1) di atas terlihat bahwa nilai parsial autokorelasi turun secara eksponensial mendekati nol.

Berdasarkan analisis ACF dan PACF dapat diperoleh karakteristik dari model MA (1) sebagai berikut.

1. Nilai autokorelasi pada *lag* 1 signifikan berbeda dengan nol.
2. Autokorelasi parsial turun secara eksponensial.

4.4.*Moving average* orde 2**, MA(2) atau ARIMA (0,0,2)**

Model *Moving Average* orde 2, dapat ditulis:

$$Z_t = (1 - \theta_1 B - \theta_2 B)\, a_t$$

Model MA(2) menandakan bahwa orde dari $p = 0$, $d = 0$ dan $q = 2$.

Karakteristik dari model MA(2) sebagai berikut.

Berdasarkan analisis ACF dan PACF dapat diperoleh karakteristik dari model MA (2) sebagai berikut.

1. Nilai autokorelasi pada *lag* 1 dan *lag 2* signifikan berbeda dengan nol.
2. Autokorelasi parsial turun secara eksponensial.

C. Studi Kasus

Studi Kasus 1. Berikut ini adalah contoh studi kasus untuk data stasioner

Tabel 1. Data penjualan harian selama 149 hari (data dibaca dari kiri ke kanan)

1008	1015	1006	1017	1015	1006	1013
1009	1011	1009	995	1024	1008	999
997	999	1004	1001	1003	1018	1026
1015	1019	1034	1033	1021	1017	1008
1009	984	964	964	976	985	997
1008	999	1003	1024	1029	1036	1049
1039	1040	1019	1007	1008	1020	1022
1023	1031	1011	1019	1009	1005	1012
1017	1000	997	983	984	992	967
966	970	982	995	981	990	1003
1005	1016	1028	1003	993	995	1003
1003	1013	1020	1019	1014	1014	1018
998	996	994	983	994	992	997
990	993	988	1004	995	995	1011
991	994	991	999	995	996	1002
1010	1016	1017	1029	1036	1024	1022
1021	1019	1009	1021	1025	1020	1021
1030	1015	1014	1016	995	1002	993
1018	1008	1017	996	1005	1014	1003
1000	1018	1002	997	1002	1007	990
994	1005	1012	1012	1028	1002	998
1029	1026					

Berdasarkan data tersebut,

a. Tentukan model ARIMA terbaik yang sesuai untuk data tersebut!

b. Lakukan peramalan untuk menentukan nilai-nilai 12 tahap ke depan!

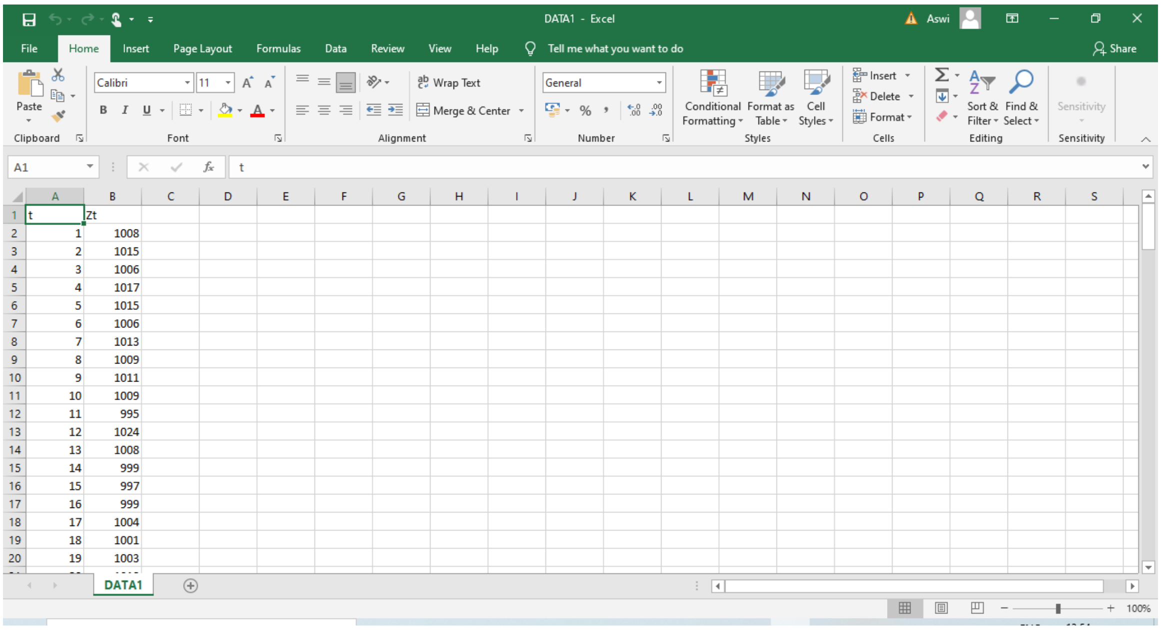
Penyelesaian:

Tahapan analisis beserta hasil outputnya dengan menggunakan software R adalah sebagai berikut:

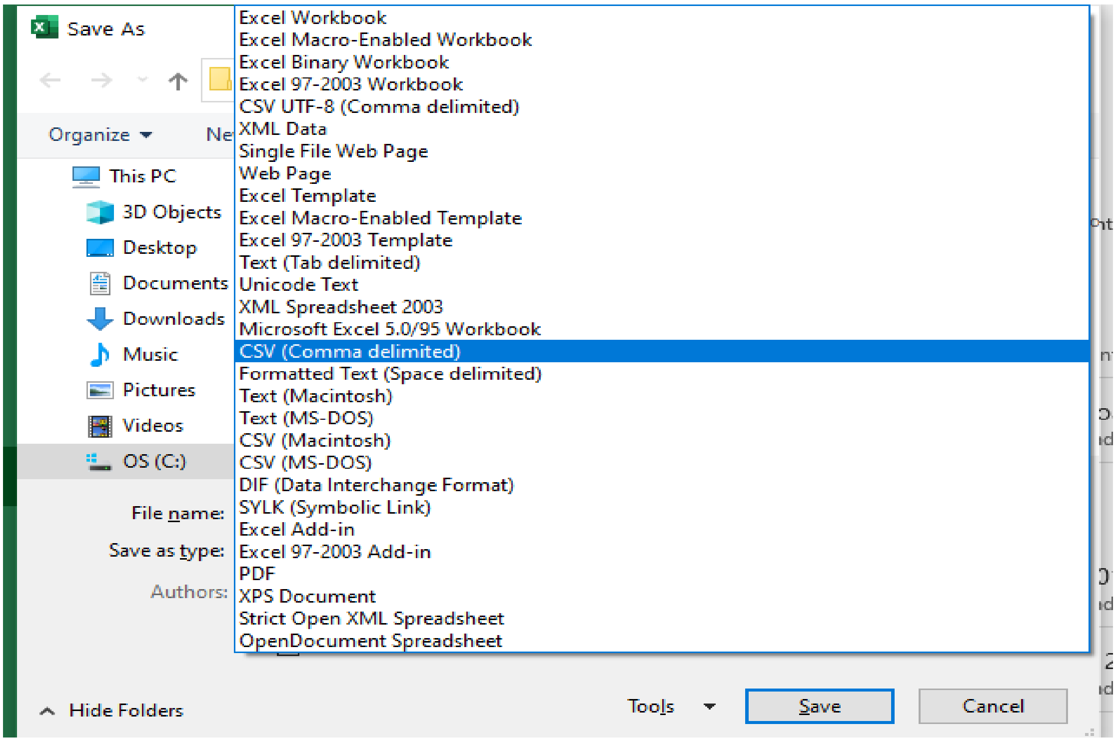
Input Data

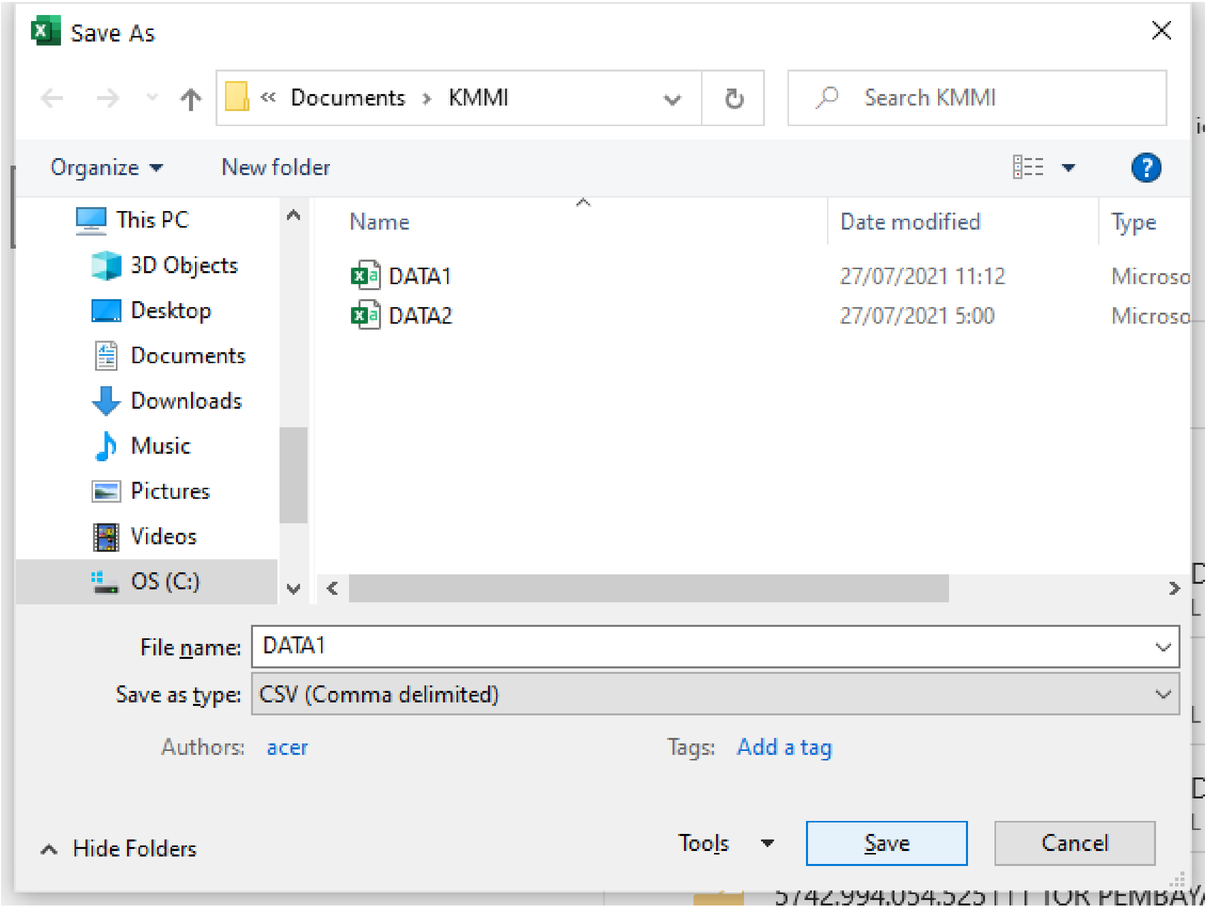
Berikut ini diberikan langkah-langkah analisis deret waktu dengan menggunakan R Studio.

1. Langkah awal adalah input data pada Table 1 di Excel dengan data terurut sesuai waktu dan memanjang ke bawah.



2. Simpan file ke dalam format *.csv misalnya dengan nama DATA1. csv dengan Langkah sebagai berikut:


Save As
Organize
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
OS (C:)
File name:
Save as type:
Authors:
Excel Workbook
Excel Macro-Enabled Workbook
Excel Binary Workbook
Excel 97-2003 Workbook
CSV UTF-8 (Comma delimited)
XML Data
Single File Web Page
Web Page
Excel Template
Excel Macro-Enabled Template
Excel 97-2003 Template
Text (Tab delimited)
Unicode Text
XML Spreadsheet 2003
Microsoft Excel 5.0/95 Workbook
CSV (Comma delimited)
Formatted Text (Space delimited)
Text (Macintosh)
Text (MS-DOS)
CSV (Macintosh)
CSV (MS-DOS)
DIF (Data Interchange Format)
SYLK (Symbolic Link)
Excel Add-in
Excel 97-2003 Add-in
PDF
XPS Document
Strict Open XML Spreadsheet
OpenDocument Spreadsheet
Hide Folders
Tools
Save
Cancel


Save As
Documents › KMMI
Search KMMI
Organize
New folder
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
OS (C:)
Name
Date modified
Type
DATA1
27/07/2021 11:12
DATA2
27/07/2021 5:00
File name: DATA1
Save as type: CSV (Comma delimited)
Authors: acer
Tags: Add a tag
Hide Folders
Tools
Save
Cancel

3. Buka R studio, lalu panggil data tersebut dengan mengetik script (tergantung dimana Anda menyimpan data Anda), misalnya:

   setwd("C:\\Users\\Aswi\\Documents\\KMMI")

   Untuk kemudahan, kita bisa mengcopy directory tempat kita menyimpan data di Excel,

   C:\Users\Aswi\Documents\KMMI (dicopy dari Excel)

   - Cara mengimport data csv ke R studio dengan menggunakan sintaks berikut:

   setwd("C:\\Users\\Aswi\\Documents\\KMMI") atau

   setwd("C:/Users/Aswi/Documents/KMMI")

   Perhatikan: gunakan tanda / atau \\ bukan tanda / untuk menentukan direktori.

4. Membaca data dengan menulis script

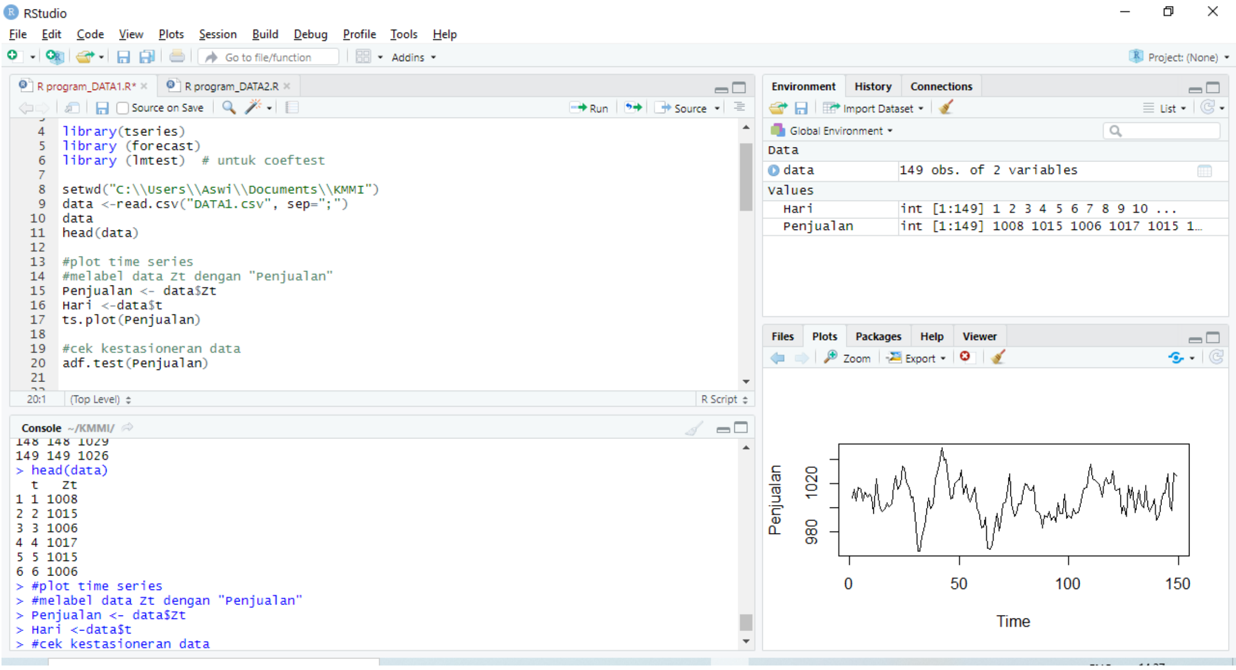
   data <-read.csv("DATA1.csv", sep=";")

   data

5. Memeriksa apakah data sudah terbaca di R dengan menuliskan script “data”, lalu run.

   Untuk melihat sebagian data awal dapat digunakan perintah “head(data) sebagai berikut:

```
> head(data)
  t   Zt
1 1 1008
2 2 1015
3 3 1006
4 4 1017
5 5 1015
6 6 1006
```



Tahap Identifikasi

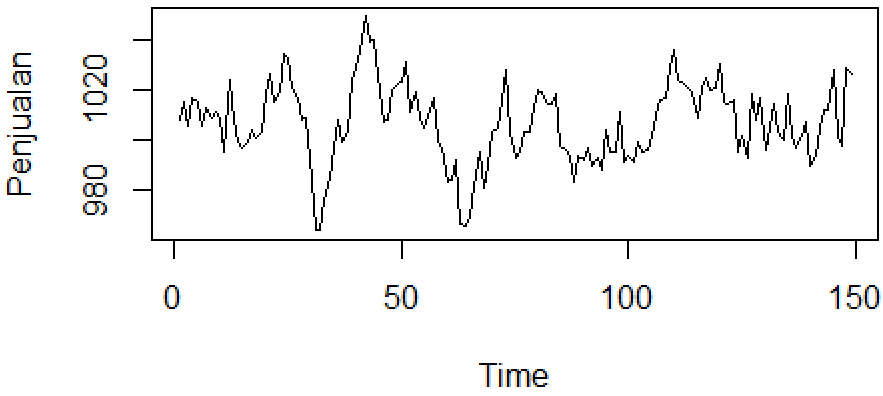
Langkah pertama yang dilakukan dalam pemodelan deret waktu adalah tahap identifikasi. Pada tahap identifikasi ini, hal yang dilakukan adalah memeriksa kestasioneran data baik dalam mean maupun dalam variansi. Kestasioneran data dalam rata-rata dapat diperiksa melalui plot time series secara visual atau melalui uji *Augmented Dickey-Fuller*.

Karena data terdiri dari dua variable yaitu t sebagai variable waktu dan Z_t sebagai data deret waktu penjualan, maka untuk kemudahan, Anda dapat melabel data$Zt dengan nama penjualan:

```
Penjualan <- data$Zt.
```

Sehingga perintah untuk plot time series adalah:

```
ts.plot (Penjualan)
```



1. Pengujian kestasioneran data melalui "*Augmented Dickey-Fuller Test*" dengan perintah:

```
> adf.test(Penjualan)
```

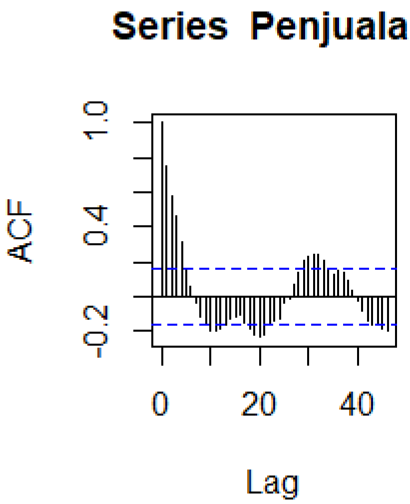
```
> adf.test(Penjualan)

        Augmented Dickey-Fuller Test

data:  Penjualan
Dickey-Fuller = -3.8973, Lag order = 5, p-value = 0.01619
alternative hypothesis: stationary
```

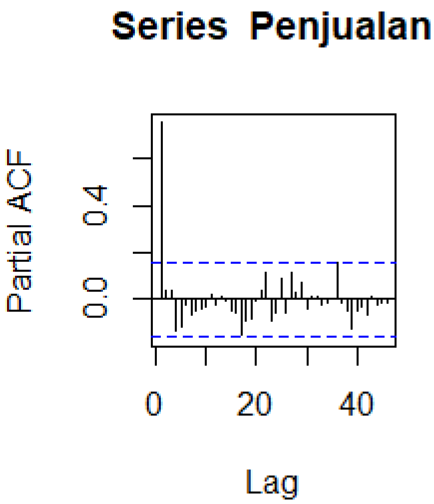
Dari hasil pengujian "*Augmented Dickey-Fuller Test*" diperoleh nilai p-value = 0,01619 yang berari bahwa dipotesis H0 ditolak, alternative hypothesis: stasionary diterima. Ini berarti bahwa data tersebut sudah stasioner dalam rata-rata.

2. Selanjutnya adalah melihat plot ACF dan PACF melalui perintah

```
par(mfrow=c(1,2))
acf(Penjualan,lag.max=46)
pacf(Penjualan,lag.max=46)
```

par(mfrow=c(1,2)) merupakan perintah untuk membuat plot menjadi 1 baris 2 kolom





Berdasarkan plot ACF dan plot PACF tersebut, dapat dilihat bahwa plot ACF turun secara eksponensial (*dies down*), dan plot PACF nya terpotong setelah lag 1 (*cut off after lag 1*). Berdasarkan teori yang telah dibahas, model dugaan sementara untuk data tersebut adalah ARIMA (1,0,0).

3. Memodelkan data dengan menggunakan ARIMA(1, 0, 0) melalui perintah:

```
> #dugaan model ARIMA sementara yaitu ARIMA(1,0,0)
> fit1 <-arima (Penjualan, order=c(1,0,0), method="ML")
> fit1

Call:
arima(x = Penjualan, order = c(1, 0, 0), method = "ML")

Coefficients:
         ar1  intercept
      0.7519  1007.3008
s.e.  0.0535     3.3302

sigma^2 estimated as 105.7:  log likelihood = -559.05,  aic = 
1124.11
```

4. Melakukan diagnostic checking yaitu uji kesignifikanan parameter sebagai berikut:

```
> #Diagnostic Checking
> #1. Uji kesignifikanan parameter: Uji t
> coeftest(fit1)

z test of coefficients:

            Estimate Std. Error z value  Pr(>|z|)    
ar1       7.5194e-01 5.3481e-02   14.06 < 2.2e-16 ***
intercept 1.0073e+03 3.3302e+00  302.48 < 2.2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Berdasarkan hasil R, taksiran parameter model AR(1) adalah signifikan berbeda dari nol. Hal ini dapat dilihat pada nilai Pr (>|z|) < 2,2e-16 < α = 0,05. Pengujian signifikan jika Pr (>|z|) < α dan pengujian tidak signifikan jika Pr (>|z|) ≥α.

5. Pengujian residual White noise melalui *L-Jung Box test*

```
> #2. pengujian residual apakah white noise
> Box.test(fit1$residuals, type="Ljung")

	Box-Ljung test

data:  fit1$residuals
X-squared = 0.13643, df = 1, p-value = 0.7119
```

Dengan melihat nilai p-value, kriteria kesimpulannya adalah menolak H0 yang menyatakan residual white noise jika p-value < α dan menerima H0 jika p-value ≥ α. Berdasarkan hasil output pengujian residual Ljung Box diperoleh p-value = 0,7119 > α = 0,05 yang berarti bahwa gagal tolak H0. Jadi dapat disimpulkan bahwa residual white noise.

6. Pengujian residual berdistribusi normal

Pengujian apakah residual berdistribusi normal atau tidak dapat dilakukan melalui beberapa cara yaitu: uji Shapiro-Wilk, uji Shapiro-Francia, uji Anderson-Darling, atau uji Lilliefors (Kolmogorov-Smirnov). Program R dan hasilnya diberikan sebagai berikut:

```
> shapiro.test(fit1$residuals)

	Shapiro-Wilk normality test

data:  fit1$residuals
W = 0.99342, p-value = 0.7323

> #Shapiro-Francia normality test
> sf.test(fit1$residuals)

	Shapiro-Francia normality test

data:  fit1$residuals
W = 0.99302, p-value = 0.5974

> #Anderson-Darling normality test
> ad.test(fit1$residuals)

	Anderson-Darling normality test

data:  fit1$residuals
```

```
A = 0.39783, p-value = 0.3628

> #Lilliefors (Kolmogorov-Smirnov) normality test
> lillie.test(fit1$residuals)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  fit1$residuals
D = 0.063826, p-value = 0.1447
```

Berdasarkan hasil output pengujian normalitas diperoleh nilai p-value > α = 0,05 yang berarti bahwa gagal tolak H0. Jadi dapat disimpulkan bahwa residual memenuhi asumsi distribusi normal.

7. Tahap peramalan dilakukan ketika parameter dalam model signifikan dan residual white noise dan berdistribusi normal. Perintah untuk tahap peramalan dah hasil output melalui software R adalah sebagai berikut.

```
> #Forecasting untuk 12 tahap ke depan
> forecasting <-forecast(Penjualan, model=fit1, h=12)
> forecasting
    Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
150       1021.362 1008.1858 1034.537 1001.2110 1041.512
151       1017.874 1001.3886 1034.359  992.6620 1043.085
152       1015.251  997.1608 1033.341  987.5844 1042.918
153       1013.279  994.3412 1032.217  984.3162 1042.242
154       1011.796  992.3955 1031.197  982.1254 1041.467
155       1010.681  991.0236 1030.338  980.6176 1040.744
156       1009.842  990.0413 1029.644  979.5592 1040.126
157       1009.212  989.3300 1029.094  978.8051 1039.619
158       1008.738  988.8104 1028.665  978.2614 1039.214
159       1008.381  988.4282 1028.335  977.8656 1038.897
160       1008.113  988.1456 1028.081  977.5753 1038.651
161       1007.912  987.9358 1027.888  977.3612 1038.462
```

Jadi hasil ramalan penjualan hari ke 150 adalah 1021.

Program R keseluruhan untuk pemodelan Studi Kasus 1 div=berikan sebagai berikut:

```
#set working directory
library(tseries)
library (forecast)
library (lmtest)  # untuk coeftest
library(normtest) #untuk normality test
library(nortest)
#set working directory
setwd("C:\\Users\\Aswi\\Documents\\KMMI")
```

```
data <-read.csv("DATA1.csv", sep=";")
data
head(data)

#plot time series
#melabel data Zt dengan "Penjualan"
Penjualan <- data$Zt
Hari <-data$t
ts.plot(Penjualan)

#cek kestasioneran data
adf.test(Penjualan)

#plot Autocorrelation Function (ACF) dan partial (PACF)
acf(Penjualan,lag.max=46)

par(mfrow=c(1,2))
acf(Penjualan,lag.max=46)
pacf(Penjualan,lag.max=46)

#dugaan model ARIMA sementara yaitu ARIMA(1,0,0)
fit1 <-arima (Penjualan, order=c(1,0,0), method="ML")
fit1

#Diagnostic Checking
#1. Uji kesignifikanan parameter: Uji t
coeftest(fit1)

#2. pengujian residual apakah white noise
Box.test(fit1$residuals, type="Ljung")

#3.pengujian residual apakah berdistribusi normal
```

```
shapiro.test(fit1$residuals)

#Shapiro-Francia normality test
sf.test(fit1$residuals)

#Anderson-Darling normality test
ad.test(fit1$residuals)

#Lilliefors (Kolmogorov-Smirnov) normality test
lillie.test(fit1$residuals)

#Forecasting untuk 12 tahap ke depan
forecasting <-forecast(Penjualan, model=fit1, h=12)
forecasting

plot(forecasting, main="Plot Hasil Peramalan")
```

Studi Kasus 2. Contoh studi kasus dengan data non-stasioner

Berikut ini adalah data deret waktu tentang jumlah penumpang kereta api di suatu kota tertentu yang dicatat tiap bulan selama 102 bulan. (Data dibaca dari kiri ke kanan).

639	643	640	653	667	667	663
654	649	651	659	672	670	675
692	702	706	710	722	729	740
755	763	788	818	826	821	819
827	848	881	879	878	878	868
856	844	824	820	819	813	815
822	818	815	792	769	775	771
773	780	779	774	772	775	770
766	771	773	772	767	775	777
777	776	779	787	790	791	792
802	799	792	780	790	799	810
814	828	862	874	892	872	869
870	859	857	870	867	856	854

862	861	855	846	847	845	838
828	823	814	812			

Berdasarkan data tersebut,

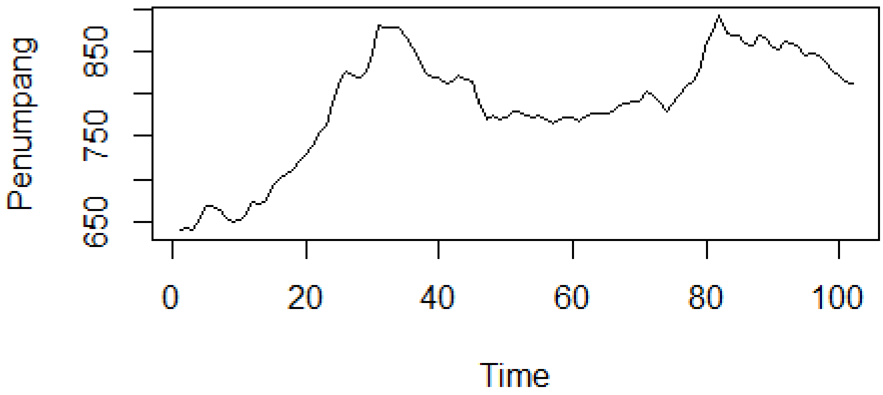
a. Tentukan model ARIMA terbaik yang sesuai untuk data tersebut!
b. Lakukan peramalan untuk menentukan nilai-nilai tiga tahap ke depan!

Penyelesaian:

Langkah input data adalah sama dengan studi kasus 1.

Tahapan pemodelan data jumlah penumpang kereta api adalah sebagai berikut:

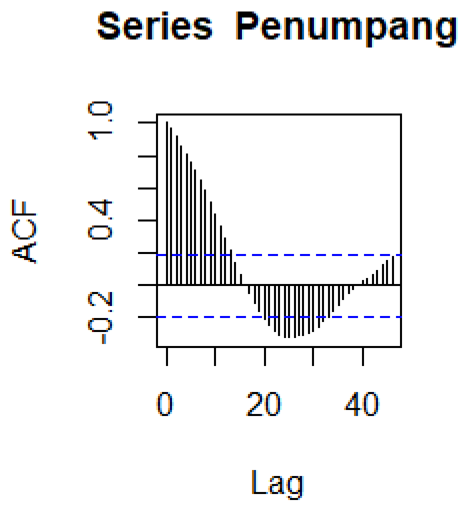
1. Plot time series



Berdasarkan plot time series data penumpang kereta api dapat disimpulkan bahwa data belum stasioner dalam rata-rata.

2. Plot ACF

Salah satu indikasi yang juga dapat digunakan untuk melihat apakah data sudah stasioner atau belum adalah melalui plot ACF nya. Jika plot ACF cenderung turun secara lambat, maka indikasi data belum stasioner dalam rata-rata.



Berdasarkan plot ACF tersebut, dapat disimpulkan bahwa data belum stsioner dalam rata-rata.

3. *Augmented Dickey-Fuler Test*

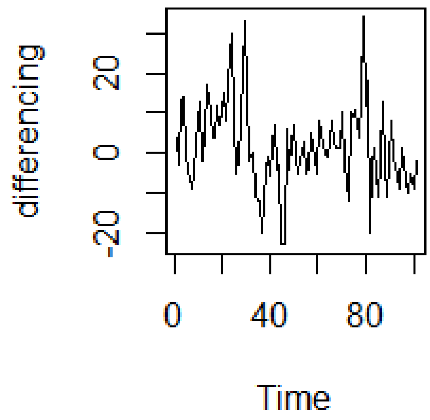
```
> adf.test(Penumpang)

	Augmented Dickey-Fuller Test

data:  Penumpang
Dickey-Fuller = -1.7806, Lag order = 4, p-value = 0.6676
alternative hypothesis: stationary
```

Dari hasil pengujian "*Augmented Dickey-Fuller Test*" diperoleh nilai p-value = 0,6676 yang berari bahwa hipotesis H0 gagal ditolak. Ini berarti bahwa data tersebut belum stasioner dalam rata-rata.

4. Differencing

```
> differencing <- diff(Penumpang, differences = 1)
> differencing
 [1]   4  -3  13  14   0  -4  -9  -5   2   8  13  -2   5  17  10   4   4  12   7  11  15
[22]   8  25  30   8  -5  -2   8  21  33  -2  -1   0 -10 -12 -12 -20  -4  -1  -6   2   7
[43]  -4  -3 -23 -23   6  -4   2   7  -1  -5  -2   3  -5  -4   5   2  -1  -5   8   2   0
[64]  -1   3   8   3   1   1  10  -3  -7 -12  10   9  11   4  14  34  12  18 -20  -3   1
[85] -11  -2  13  -3 -11  -2   8  -1  -6  -9   1  -2  -7 -10  -5  -9  -2
```

5. Plot time series untuk data hasil differencing



Berdasarkan plot timeseries hasil differencing dapat disimpulkan bahwa data masih belum stasioner dalam rata-rata, sehingga perlu dilakukan proses differencing lagi.

```
> adf.test(differencing)

        Augmented Dickey-Fuller Test

data:  differencing
Dickey-Fuller = -3.3317, Lag order = 4, p-value = 0.06972
alternative hypothesis: stationary
```

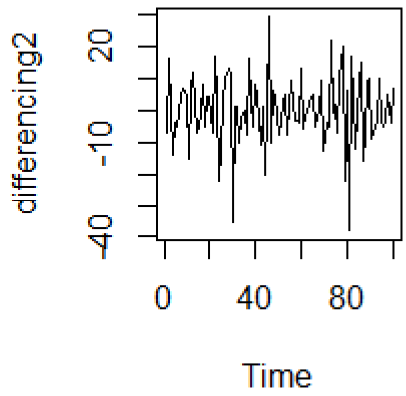
Dari hasil pengujian "*Augmented Dickey-Fuller Test*" diperoleh nilai p-value = 0,0672 yang berari bahwa hipotesis H0 gagal ditolak. Ini berarti bahwa data tersebut belum stasioner dalam rata-rata. Oleh karena itu, dapat dilakukan differencing lagi dengan melakukan differncing 1 dari hasil differencing sebagai berikut:

```
> differencing2 <- diff(differencing, differences = 1)
> differencing2
 [1]  -7  16   1 -14  -4  -5   4   7   6   5 -15   7  12  -7  -6   0   8  -5   4   4  -7
[22]  17   5 -22 -13   3  10  13  12 -35   1   1 -10  -2   0  -8  16   3  -5   8   5 -11
[43]   1 -20   0  29 -10   6   5  -8  -4   3   5  -8   1   9  -3  -3  -4  13  -6  -2  -1
[64]   4   5  -5  -2   0   9 -13  -4  -5  22  -1   2  -7  10  20 -22   6 -38  17   4 -12
[85]   9  15 -16  -8   9  10  -9  -5  -3  10  -3  -5  -3   5  -4   7
```

Alternative kedua adalah dengan langsung menulis differences=2, tetapi menggunakan data awal sebagai berikut:

```
> #differencing 2 non musiman (cara lain)
> differencing2 <- diff(Penumpang, differences = 2)
> differencing2
 [1]  -7  16   1 -14  -4  -5   4   7   6   5 -15   7  12  -7  -6   0   8  -5   4   4  -7
[22]  17   5 -22 -13   3  10  13  12 -35   1   1 -10  -2   0  -8  16   3  -5   8   5 -11
[43]   1 -20   0  29 -10   6   5  -8  -4   3   5  -8   1   9  -3  -3  -4  13  -6  -2  -1
[64]   4   5  -5  -2   0   9 -13  -4  -5  22  -1   2  -7  10  20 -22   6 -38  17   4 -12
[85]   9  15 -16  -8   9  10  -9  -5  -3  10  -3  -5  -3   5  -4   7
```

6. Time series plot hasil differencing I = 2

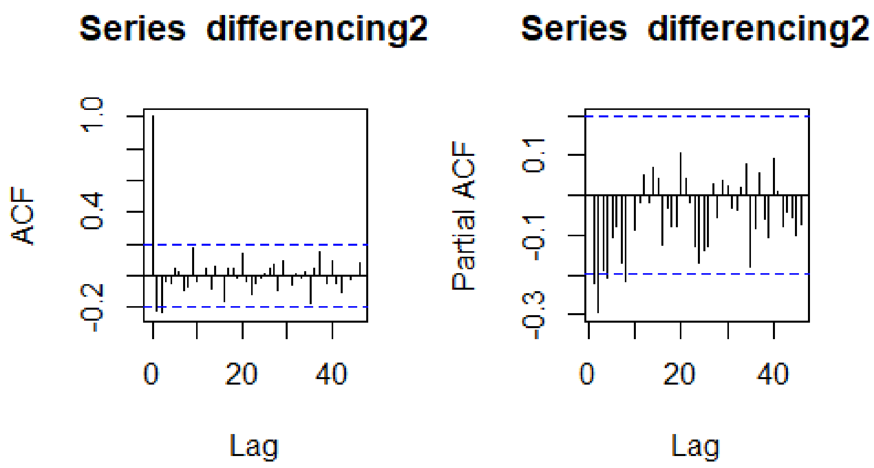


7. Pengujian dengan ADT

```
> adf.test(differencing2)

        Augmented Dickey-Fuller Test

data:  differencing2
Dickey-Fuller = -6.6191, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Dari hasil pengujian "*Augmented Dickey-Fuller Test*" diperoleh nilai p-value = 0,01 yang berari bahwa hipotesis H0 ditolak. Ini berarti bahwa data tersebut sudah stasioner dalam rata-rata.

8. Plot ACF dan PACF dari data hasil differencing 2



Berdasarkan plot ACF dan plot PACF tersebut, dapat dilihat bahwa plot ACF nya terpotong setelah lag 2 (*cut off after lag 2*), dan plot PACF nya turun secara eksponensial (*dies down*). Berdasarkan teori yang telah dibahas, model dugaan sementara untuk data tersebut adalah ARIMA (0, 2, 2).

8. Memodelkan data dengan menggunakan ARIMA(0, 2, 2) melalui perintah sebagai berikut:

```
> fit1 <-arima (Penumpang, order=c(0,2,2), method="ML")
> fit1

Call:
arima(x = Penumpang, order = c(0, 2, 2), method = "ML")

Coefficients:
        ma1      ma2
    -0.4749  -0.3614
s.e.  0.0949   0.1025
```

```
sigma^2 estimated as 86.21:  log likelihood = -365.27,  aic = 736.55
```

9. Melakukan diagnostic checking yaitu uji kesignifikanan parameter sebagai berikut:

```
> #Diagnostic Checking
> #1. Uji kesignifikanan parameter: Uji t
> coeftest(fit1)
z test of coefficients:
         Estimate  Std. Error   z value    Pr(>|z|)
ma1 -0.474874   0.094858   -5.0062    5.553e-07 ***
ma2 -0.361393   0.102471   -3.5268     0.0004207 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Berdasarkan hasil R, taksiran parameter model MA(1) dan MA(2) adalah signifikan berbeda dari nol. Hal ini dapat dilihat pada nilai Pr (>|z|) =5,5e-07 < α = 0,05, dan Pr (>|z|) =0.0004207 < α = 0,05.

10. Pengujian residual White noise melalui *L-Jung Box test*

```
> #2. pengujian residual apakah white noise
> Box.test(fit1$residuals, type="Ljung")

        Box-Ljung test
data:  fit1$residuals
X-squared = 0.0058277, df = 1, p-value = 0.9391
```

Berdasarkan hasil output pengujian residual Ljung Box diperoleh p-value = 0,9391 > α = 0,05 yang berarti bahwa gagal tolak H0. Jadi dapat disimpulkan bahwa residual white noise.

11. Pengujian residual berdistribusi normal

Pengujian apakah residual berdistribusi normal atau tidak dapat dilakukan melalui beberapa cara yaitu: uji Shapiro-Wilk, uji Shapiro-Francia, uji Anderson-Darling, atau uji Lilliefors (Kolmogorov-Smirnov). Program R dan hasilnya diberikan sebagai berikut:

```
> #3. Penujian residual apakah berdistribusi normal
> shapiro.test(fit1$residuals)

        Shapiro-Wilk normality test

data:  fit1$residuals
W = 0.98258, p-value = 0.1992

> #Shapiro-Francia normality test
> sf.test(fit1$residuals)

        Shapiro-Francia normality test

data:  fit1$residuals
W = 0.97719, p-value = 0.06975

> #Anderson-Darling normality test
> ad.test(fit1$residuals)

        Anderson-Darling normality test

data:  fit1$residuals
A = 0.36995, p-value = 0.4193

> #Lilliefors (Kolmogorov-Smirnov) normality test
> lillie.test(fit1$residuals)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  fit1$residuals
D = 0.057109, p-value = 0.5695
```

Berdasarkan hasil output pengujian normalitas diperoleh nilai p-value > α = 0,05 yang berarti bahwa gagal tolak H0. Jadi dapat disimpulkan bahwa residual memenuhi asumsi distribusi normal.

12. Forecasting untuk 12 tahap ke depan

```
> #Forecasting untuk 12 tahap ke depan
> forecasting <-forecast(Penumpang, model=fit1, h=12)
```

```
> forecasting
 Point   Forecast   Lo 80     Hi 80     Lo 95     Hi 95
103      810.3680  798.4691  822.2669  792.1702  828.5658
104      807.2774  785.5769  828.9778  774.0894  840.4653
105      804.1867  774.6107  833.7628  758.9541  849.4194
106      801.0961  764.2088  837.9834  744.6818  857.5103
107      798.0054  754.0021  842.0087  730.7082  865.3027
108      794.9148  743.8345  845.9950  716.7943  873.0352
109      791.8241  733.6265  850.0217  702.8185  880.8297
110      788.7335  723.3332  854.1337  688.7124  888.7546
111      785.6428  712.9277  858.3579  674.4347  896.8510
112      782.5522  702.3934  862.7109  659.9600  905.1444
113      779.4615  691.7199  867.2031  645.2723  913.6508
114      776.3709  680.9005  871.8412  630.3615  922.3802
```

Jadi hasil ramalan penumpang bulan ke 103 adalah 810

Program R keseluruhan untuk pemodelan Studi Kasus 2 diberikan sebagai berikut:

```
library(tseries)
library (forecast)
library (lmtest)  # untuk coeftest
library(normtest) #untuk normality test
library(nortest)
#set working directory
setwd("C:\\Users\\Aswi\\Documents\\KMMI")
data <-read.csv("DATA2.csv", sep=";")
data
head(data)

#plot time series
#melabel data Zt dengan "penumpang"
Penumpang <- data$Zt
```

```
Bulan <-data$t
ts.plot(Penumpang)

#cek kestasioneran data
adf.test(Penumpang)

#plot Autocorrelation Function (ACF) dan partial (PACF)

par(mfrow=c(1,2))
acf(Penumpang,lag.max=46)
pacf(Penumpang,lag.max=46)

#differencing 1 non musiman
differencing <- diff(Penumpang, differences = 1)
differencing

#differencing 2 non musiman
differencing2 <- diff(differencing, differences = 1)
differencing2

#differencing 2 non musiman (cara lain)
differencing2 <- diff(Penumpang, differences = 2)
differencing2

ts.plot(differencing2)

#cek kestasioneran data
adf.test(differencing2)

par(mfrow=c(1,2))
acf(differencing2,lag.max=46)
pacf(differencing2,lag.max=46)
```

```
#dugaan model ARIMA sementara yaitu ARIMA(0,2,2)
fit1 <-arima (Penumpang, order=c(0,2,2), method="ML")
fit1

#Diagnostic Checking
#1. Uji kesignifikanan parameter: Uji t
coeftest(fit1)

#2. pengujian residual apakah white noise
Box.test(fit1$residuals, type="Ljung")

#3. Penujian residual apakah berdistribusi normal
shapiro.test(fit1$residuals)

#Shapiro-Francia normality test
sf.test(fit1$residuals)

#Anderson-Darling normality test
ad.test(fit1$residuals)

#Lilliefors (Kolmogorov-Smirnov) normality test
lillie.test(fit1$residuals)

#Forecasting untuk 12 tahap ke depan
forecasting <-forecast(Penumpang, model=fit1, h=12)
forecasting

plot(forecasting, main="Plot Hasil Peramalan")
```

D. Latihan Mandiri

Berikut ini adalah data deret waktu tentang besarnya "penjualan dari suatu produk tertentu" yang diamati setiap minggu selama 90 minggu pengamatan (data dibaca dari kiri ke kanan)

235,000	239,000	244,090	252,731	264,377	277,934
286,687	295,629	310,444	325,112	336,291	344,459
355,399	367,691	384,003	398,042	412,969	422,901
434,960	445,853	455,929	465,584	477,894	491,408
507,712	517,237	524,349	532,104	538,097	544,948
551,925	557,929	564,285	572,164	582,926	595,295
607,028	617,541	622,941	633,436	647,371	658,230
670,777	685,457	690,992	693,557	700,675	712,710
726,513	736,429	743,203	751,227	764,265	777,852
791,070	805,844	815,122	822,905	830,663	839,600
846,962	853,830	860,840	871,075	877,792	881,143
884,226	890,208	894,966	901,288	913,138	922,511
930,786	941,306	950,305	952,373	960,042	968,100
972,477	977,408	977,602	979,505	982,934	985,833
991,350	996,291	1003,100	1010,320	1018,420	1029,480

a. Dengan menggunakan program R tentukan model ARIMA terbaik untuk data tersebut! Lampirkan hasil komputer yang diperoleh untuk kesesuaian modelnya!

b. Berdasarkan hasil tersebut, tuliskan secara lengkap model ARIMA Box-Jenkins untuk data penjualan tersebut!

c. Ujilah apakah taksiran parameter model yang diperoleh tersebut signifikan berbeda dari nol dengan tingkat keyakinan 95%! Apa kesimpulan anda?

d. Lakukan pemeriksaan diagnostik (*diagnostic checking*) untuk menguji apakah sisa sudah memenuhi syarat cukup (*residual white noise*)! Jelaskan kesimpulan anda!

e. Carilah nilai sisa untuk data pada t = 3 atau ($Z_3 - \hat{Z}_3$)!

f. Carilah nilai ramalan untuk $\hat{Z}_{91}(90)$, $\hat{Z}_{92}(90)$ dan $\hat{Z}_{93}(90)$, yaitu nilai ramalan untuk Z_{91}, Z_{92}, dan Z_{93} yang dibuat berdasarkan pengamatan pada waktu t = 90!

E. Rangkuman

Modul ini telah memperkenalkan analisis deret waktu untuk data non musiman baik yang stasioner maupun non stasioner yang dapat digunakan untuk meramalkan data beberapa tahap ke depan. Tahapan metodologi Box-Jenkins juga sudah diperkenalkan. Modul ini juga telah dilengkapi dengan beberapa studi kasus beserta tahapan pengolahan datanya dengan menggunakan software R.

F. Daftar Pustaka

1. Aswi dan Sukarna. 2017. Cetakan Kedua. *Analisis Deret Waktu*. Andira Publisher: Makassar.
2. Bowerman, B.L. & O'Connell, R.T., 1993. *Forecasting and Time Series : An Applied Approach*, 3rd Edition, Miami University, Ohio : Duxbury Press.
3. Box, G.E.P & Cox, D.R., 1964. An Analysis of Transformations. *Journal Royal Stat. Soc.*, *ser. B.*, 26, *hal.* 211 – 252.
4. Box, G.E.P & Jenkins, G.M., 1976. *Time Series Analysis Forecasting and Control*, 2nd Edition, San Francisco: Holden-Day.
5. Cryer, J.D., 1986. *Time Series Analysis*, Massachusetts : PWS Publishers.
6. Hamilton, J.D., 1994. *Time Series Analysis*, New Jersey : Princeton University Press.
7. Hanke, J.E., Reitsch, A.G., & Wichern, D.W., 2001. *Business Forecasting,* 7th Edition, Singapore: Prentice-Hall.
8. Makridakis, S., Wheelwright, S.C., & McGee, V.E., 1998. *Forecasting: Methods and Applications*, 2nd Edition, John Wiley & Sons, Inc. (alih bahasa: Hari Sumintro, 1999, *Metode dan Aplikasi Peramalan*, Edisi ke-2, Jakarta: Binarupa Aksara.)
9. Wei, W.W.S., 1994. *Time Series Analysis: Univariate and Multivariate Methods*, California: Addison-Wesley Publishing Company.