Nama **: Andi Illa Erviani Nensi**
Nim **: 200102502002**
Kelas **: DS02**

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!

Jawaban:

Langkah-langkah untuk menganalisis deret waktu di R studio:

Install packages berikut:

```
install.packages('tidyverse')
install.packages('tseries')
install.packages('lmtest')
install.packages('normtest')
install.packages('nortest')
```

Selanjutnya, panggil packages berikut dengan syntax:

```
library(tidyverse)
library(tseries)
library(lmtest)
library(normtest)
library(nortest)
```

Setelah memanggil packages, kita memanggil data dari exel dengen menggunakan syntax *setwd* atau data di input ke R dan data terurut kebawah sesuai dengan waktu(minggu) dan memanjang ke bawah.

```
#input data
Zt = c(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)
t = c(1:90)
```

Selanjutnya memeriksa apakah data sudah terbaca dengan benar di R dengan menuliskan script ‘data’, atau ‘data=frame.(t,Zt) kemudian di run. Dengan melihat sebagaian data awal digunakan perintah ‘head(data)’ atau melihat data terbawah dengan perintah ‘tail(data)’ sebagai berikut:

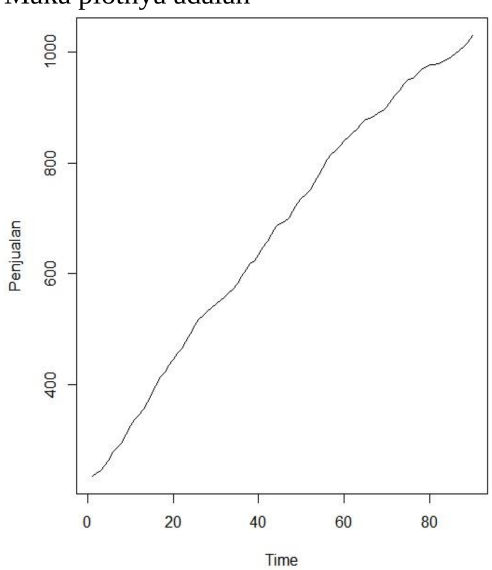
```
> head(data)
  t      Zt
1 1 235,000
2 2 239,000
3 3 244,090
4 4 252,731
5 5 264,377
6 6 277,934
```

```
> tail(data)
    t        Zt
85 85  991,350
86 86  996,291
87 87 1003,100
88 88 1010,320
89 89 1018,420
90 90 1029,480
```

Langkah berikutnya adalah tahap identifikasi dengan dilakukannya pemodelan deret waktu yaitu tahap identifikasi. Dalam tahap identifikasi dilakukan pemeriksanaan 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. Data tersebut terdiri dari dua variabel yaitu t sebagai variabel waktu ddan Zt sebagai data deret waktu penjualan. Dan berikut juga terdapat perintah untuk menampilkan plot time series yaitu 'ts.plot(Penjualan)'

```
Penjualan<- data$Zt
Minggu <- data$t
ts.plot(Penjualan)
```

Maka plotnya adalah



Selanjutnya dilakukan pengujian kestasioneran data melalui "Augmented Dickey-Fuller Test" dengan perintah:

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

```
> adf.test(Penjualan)

        Augmented Dickey-Fuller Test

data:  Penjualan
Dickey-Fuller = -0.74961, Lag order = 4, p-value = 0.9627
alternative hypothesis: stationary
```

Dari hasil pengujian "Augmented Dickey-Fuller Test" dapat diperoleh bahwa p-value = 0,9627 maka hipotesis H0 diterima atau gagal di tolak berarti H1 di tolak sehingga dilakukan altervnatif hypothesis yaitu stationary ditolak. Maka data tersebut belum stasioner dalam rata-rata.

Selanjutnya dilakukan Differencing

Differencing yang dilakukanadalah differencing 1 dengan perintah sebagai berikut:

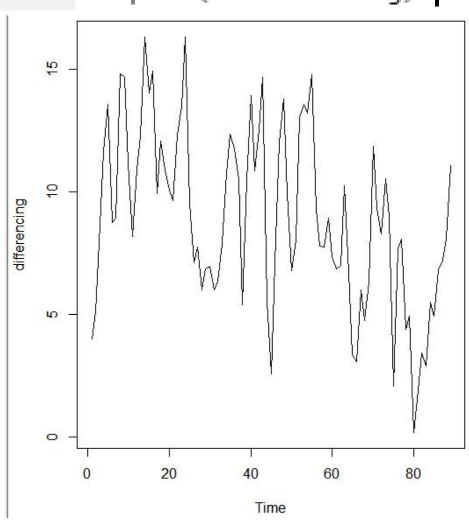
```
60  differencing <- diff(Penjualan, differences = 1) differencing
```

Maka:

```
> differencing <- diff(Penjualan, differences = 1)
> differencing
 [1]  4.000  5.090  8.641 11.646 13.557  8.753  8.942 14.815 14.668 11.179  8.168
[12] 10.940 12.292 16.312 14.039 14.927  9.932 12.059 10.893 10.076  9.655 12.310
[23] 13.514 16.304  9.525  7.112  7.755  5.993  6.851  6.977  6.004  6.356  7.879
[34] 10.762 12.369 11.733 10.513  5.400 10.495 13.935 10.859 12.547 14.680  5.535
[45]  2.565  7.118 12.035 13.803  9.916  6.774  8.024 13.038 13.587 13.218 14.774
[56]  9.278  7.783  7.758  8.937  7.362  6.868  7.010 10.235  6.717  3.351  3.083
[67]  5.982  4.758  6.322 11.850  9.373  8.275 10.520  8.999  2.068  7.669  8.058
[78]  4.377  4.931  0.194  1.903  3.429  2.899  5.517  4.941  6.809  7.220  8.100
[89] 11.060
```

Langkah berikutnya dilakukan plot time series hasil differencing, maka:

```
62  ts.plot(differencing)
```



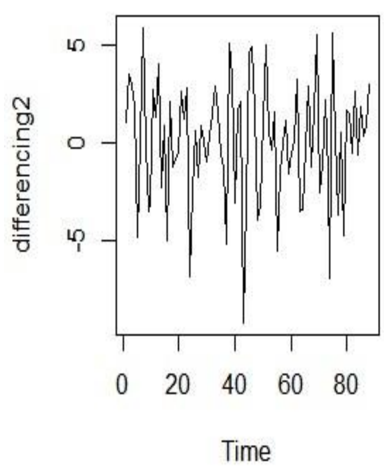
Berdasarkan plot timeseries hasil differencing sebelumnya, dapat disimpulkan bahwa data masih belum stasioner dalam rata-rata, sehingga masih perlu dilakukan proses differencing berikutnya atau differencing ke-2. Untuk lebih meyakinkan maka dilakuakn pengujian kestasioneran sebagai berikut:

```
> adf.test(differencing)

        Augmented Dickey-Fuller Test

data:  differencing
Dickey-Fuller = -3.3837, Lag order = 4, p-value = 0.06285
alternative hypothesis: stationary
```

Berdasarkan hasil pengujian “Augmented Dickey-Fuller Test” diperoleh bahwa nilai p-value = 0,06285 yang berarti bahwa hipotesis H0 gagal ditolak. Hal ini berarti bahwa data tersebut masih belum stasioner dalam rata-rata. Maka, dapat dilakukan differencing lagi dengan melakukan differncing 1 dari hasil differencing sebagai berikut:

```
> #differencing 2 non musiman
> differencing2 <- diff(differencing, differences = 1)
> differencing2
 [1]  1.090  3.551  3.005  1.911 -4.804  0.189  5.873 -0.147 -3.489 -3.011  2.772
[12]  1.352  4.020 -2.273  0.888 -4.995  2.127 -1.166 -0.817 -0.421  2.655  1.204
[23]  2.790 -6.779 -2.413  0.643 -1.762  0.858  0.126 -0.973  0.352  1.523  2.883
[34]  1.607 -0.636 -1.220 -5.113  5.095  3.440 -3.076  1.688  2.133 -9.145 -2.970
[45]  4.553  4.917  1.768 -3.887 -3.142  1.250  5.014  0.549 -0.369  1.556 -5.496
[56] -1.495 -0.025  1.179 -1.575 -0.494  0.142  3.225 -3.518 -3.366 -0.268  2.899
[67] -1.224  1.564  5.528 -2.477 -1.098  2.245 -1.521 -6.931  5.601  0.389 -3.681
[78]  0.554 -4.737  1.709  1.526 -0.530  2.618 -0.576  1.868  0.411  0.880  2.960
```

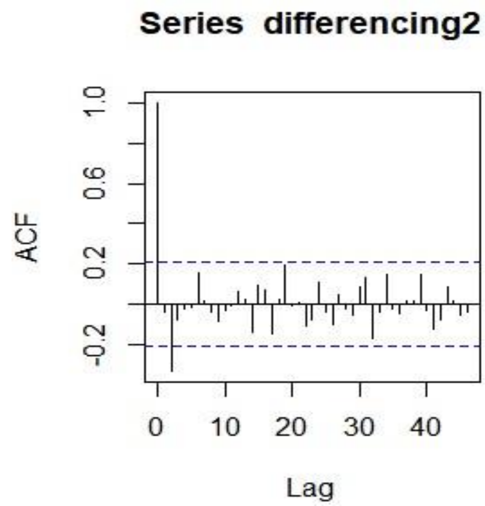
maka plot dari data hasil differencing ke-2 adalah sebagai berikut:

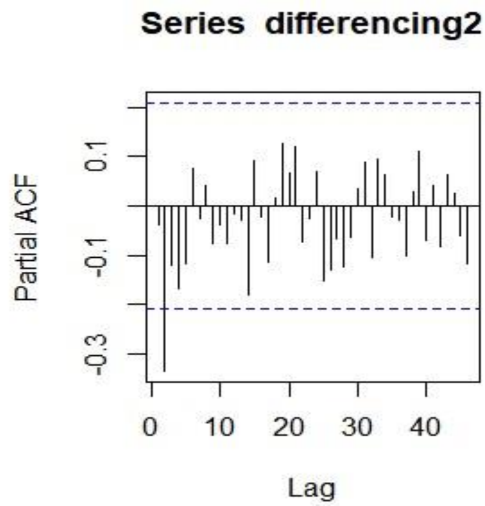


Selanjutnya dilakukan pengujian ADT sehingga diperoleh:

```
> adf.test(differencing2)

        Augmented Dickey-Fuller Test

data:  differencing2
Dickey-Fuller = -6.0598, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Dari hasil pengujian “Augmented Dickey-Fuller Test” diperoleh nilai p-value = 0,01 yang berarti bahwa hipotesis H0 ditolak atau H1 diterima. Oleh karena itu, data tersebut sudah stasioner dalam rata-rata. 6 Plot ACF dan PACF hasil differencing 2 yaitu:





Berdasarkan hasil dari plot ACF dan plot PACF tersebut, dapat dilihat bahwa plot ACF terpotong setelah lag 2 (cut off after lag 2), dan plot PACF turun secara eksponensial (dies down). Maka dapat disimpulkan bahwa model dugaan sementara untuk data tersebut adalah ARIMA (0, 2, 2).

Selanjutnya data dimodelkan dengan menggunkaan ARIMA (0,2,2) melalui perintah sebagai berikut:

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

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

Coefficients:
          ma1      ma2
      -0.2801  -0.5001
s.e.   0.1043   0.1069

sigma^2 estimated as 7.471:  log likelihood = -213.83,  aic = 433.65
```

Langkah berikutnya dilakukan diagnostic checking yaitu ujikesignifikanan parameter sebagai berikut:

```
> coeftest(fit1)

z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ma1 -0.28008    0.10431 -2.6850  0.007253 **
ma2 -0.50009    0.10689 -4.6786 2.888e-06 ***
---
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 tersebut dapat dilihat pada nilai Pr (>|z|) =0.007253 < * = 0,05, danPr (>|z|) =2.8e-06 <* = 0,05.

Langkah berikutnnya dilakukan Pengujian residual White noise melalui*L-jung Box test* sbegai berikut:

```
> Box.test(fit1$residuals, type="Ljung")

        Box-Ljung test

data:  fit1$residuals
X-squared = 0.73078, df = 1, p-value = 0.3926
```

Dari hasil output pengujian residual Ljung Box diperoleh p-value = = 0,3926 >* =0,05 yang berarti bahwa H1 di terima atau gagal tolak H0. Sehinga dapat disimpulkan bahwa residual wahite noise.

Langkah berikutnya dilakuakan pengujian berdistribusi normal yang dapat dilakuakn melalui beberapa cara yaitu: *uji Shapiro-Wilk, uji Shapiro-Francia, uji Anderson-Darling dan uji Liliefors(kolmogrov-Smirnov).* Dengan program R dan hasilnya adalah sebgai berikut:

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

        Shapiro-Wilk normality test

data:  fit1$residuals
W = 0.98343, p-value = 0.3091

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

        Shapiro-Francia normality test

data:  fit1$residuals
W = 0.98659, p-value = 0.4125

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

        Anderson-Darling normality test

data:  fit1$residuals
A = 0.38303, p-value = 0.3901

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

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  fit1$residuals
D = 0.080503, p-value = 0.1607
```

Dari hasl output pengujian normalitas sebelumnya diperoleh p-value >*=0.05 yaitu gagal tolak H0 sehingga dapat disimpulkan bahwa residual tersebut memenuhi asumsi distribusi normal.

Forecasting untuk 12 tahap ke depan:

```
> forecasting <-forecast(Penjualan, model=fit1, h=12)
> forecasting
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
 91       1038.305 1034.802 1041.808 1032.948 1043.662
 92       1045.160 1038.191 1052.129 1034.502 1055.818
 93       1052.015 1042.281 1061.748 1037.129 1066.900
 94       1058.869 1046.542 1071.197 1040.017 1077.722
 95       1065.724 1050.844 1080.605 1042.967 1088.482
 96       1072.579 1055.134 1090.024 1045.899 1099.259
 97       1079.434 1059.388 1099.480 1048.776 1110.092
 98       1086.289 1063.592 1108.986 1051.577 1121.001
 99       1093.144 1067.738 1118.549 1054.289 1131.998
100       1099.999 1071.822 1128.175 1056.906 1143.091
101       1106.853 1075.841 1137.866 1059.424 1154.283
102       1113.708 1079.795 1147.622 1061.842 1165.575
> plot(forecasting, main="Plot Hasil Peramalan")
```

Jadi dapat dilihat bahwa hasil ramalan untuk $\hat{Z}_{91}(90)$ adalah 1038.305 , $\hat{Z}_{92}(90)$ adalah 1045.160 dan$\hat{Z}_{93}$ (90) adalah 1052.015.

Sehingga, Model ARIMA terbaik untuk data tersebut adalah ARIMA (0,2,2)

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

Jawab:

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

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

Coefficients:
          ma1      ma2
      -0.2801  -0.5001
s.e.   0.1043   0.1069

sigma^2 estimated as 7.471:  log likelihood = -213.83,  aic = 433.65
```

Dari hasil tersebut, didapatkan ma1 = -0.2801 dan ma2 = -0.5001. Maka modelnya adalah: Zt = at – (-0.2801)at-1 - (-0.5001)at-2 Dengan model differencing yaitu:

$(1-B^2)Zt = (0.2801B) + (-0.5001B^2)at.$

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

Jawab: Taksiran parameter model ma(2) tersebut signifikan berbeda dari nol dengan tingkat 95%. Hal inidapat dilihat pada nilai z hitung atau nilai-p. Untuk parameter ma(1) yaitu, nilai z = -2.6850 dengan nilaiPr (>|z|) =0.007253, parameter ma(2) yaitu nilai z = -4.6786 dengan nilai Pr (>|z|) =2.8e-06. Hal tersebut dapat dilihat pada nilai Pr (>|z|) =0.007253 <* = 0,05, danPr (>|z|) =2.8e-06 <* = 0,05.

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

Jawab:

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

        Shapiro-Wilk normality test

data:  fit1$residuals
W = 0.98343, p-value = 0.3091

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

        Shapiro-Francia normality test

data:  fit1$residuals
W = 0.98659, p-value = 0.4125

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

        Anderson-Darling normality test

data:  fit1$residuals
A = 0.38303, p-value = 0.3901

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

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  fit1$residuals
D = 0.080503, p-value = 0.1607
```

Sisa tersebut sudah memenuhi syarat cukup (*residual white* noise). Dapat dilihat dari hasil output pengujian residual Ljung Box diperoleh p-value = 0,3926 >* = 0,05 yang berarti bahwa gagal tolak H0 atau tolak H1. Sehinngga disimpulkan bahwa residual tersebut white noise.

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

Jawab:

Nilaisisauntuk data pada t = 3 atau ($Z_3 - \hat{Z}_3$) yaitu:

$$MSE = \frac{1}{N}\sum_{t=3}^{N} \hat{a}^2{}_t$$

Dengan $\hat{a}^2{}_t$ = taksiran sisa pada peramalan

$\hat{a}^2$ = ($Z_3 - \hat{Z}_3$) (3)

N = Banyaknyasisa (12)

Maka:

MSE = 1/12 .3

$$MSE = 0.25$$

Jadi, MSE = 0.25

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!

Jawab: Berikut adalah Forecasting untuk 12 tahapkedepan

```
> forecasting <-forecast(Penjualan, model=fit1, h=12)
> forecasting
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
 91       1038.305 1034.802 1041.808 1032.948 1043.662
 92       1045.160 1038.191 1052.129 1034.502 1055.818
 93       1052.015 1042.281 1061.748 1037.129 1066.900
 94       1058.869 1046.542 1071.197 1040.017 1077.722
 95       1065.724 1050.844 1080.605 1042.967 1088.482
 96       1072.579 1055.134 1090.024 1045.899 1099.259
 97       1079.434 1059.388 1099.480 1048.776 1110.092
 98       1086.289 1063.592 1108.986 1051.577 1121.001
 99       1093.144 1067.738 1118.549 1054.289 1131.998
100       1099.999 1071.822 1128.175 1056.906 1143.091
101       1106.853 1075.841 1137.866 1059.424 1154.283
102       1113.708 1079.795 1147.622 1061.842 1165.575
> plot(forecasting, main="Plot Hasil Peramalan")
```

Sehingga dapat dilihat dari hasil ramalan untuk $\hat{Z}_{91}(90)$ adalah 1038.305 , $\hat{Z}_{92}(90)$ adalah 1045.160 dan $\hat{Z}_{93}(90)$ adalah 1052.015.