Scholars Journal of Physics, Mathematics and Statistics
Abbreviated Key Title: Sch. J. Phys. Math. Stat.
ISSN 2393-8056 (Print)
ISSN 2393-8064 (Online)

(An International Publisher for Academic and Scientific Resources)

On the Mathematical Report Writing, Based on Data, using Command explained Minitab

Confess Matete*
Department of Mathematics and Computer Science, Great Zimbabwe University, Masvingo, Zimbabwe

***Corresponding author**
Confess Matete

Article History
Received: 10.02.2018
Accepted: 20.02.2018
Published: 30.03.2018

DOI:
10.21276/sjpms.2018.5.2.2

Abstract: Minitab is astatistical software which is used to analyse data for quality improvement. Associated with Minitab are uniquely designed products which help professionals improve their business processes. This software is used by companies world-wide in order to deliver a quality product. Examples include Ford Motor Company, which used Minitab as a statistical tool to launch the Ford Fiesta; the Riverview Hospital Association Lean Six Sigma team used Minitab to perform data analysis to identify patient groups who were scoring lower on patient satisfaction survey question. Also, Boston Scientific's medical device manufacturer used Minitab to analyse data that would validate the pouch-sealing process used to package guide wires. In this manuscript, we intend to compile an analysed data-based report for the sales of rice in Zimbabwe. The report is based on the data supplied by Afro Foods Management requesting the likely sales of rice in Zimbabwe for the first three months of 2011. We use Minitab to analyse the data and then fit the best Autoregressive Integrated Moving Average *[ARIMA]* model and Explanatory model built using the previous data of the sales of rice. The results showed that, for short periods, ARIMA model is the best.
Keywords: Minitab, Time Series Plot, Data Differencing, ACF and PACF, ARIMA Model.

INTRODUCTION

Being a comprehensive set of powerful statistics for investigating data, Minitab has uniquely designed products which help professionals improve their business processes [1].

In Basic Statistics, Minitab accesses a complete set of statistical tools that include Descriptive Statistics, Hypothesis Tests, Confidence Intervals and Normality Tests. For Regression and Analysis of Variance (ANOVA), Minitab reveals relationships between variables and identifies important factors that affect the quality of products and services [2]. Minitab is also used in quality tools, where it determines if measurement systems are adequate and assesses how well processes meet specification limits as well as creating sampling plans. This statistical software has much application in Design of Experiments, Control Charts and Reliability and Survival. In Design of Experiments, Minitab finds settings that optimize processes using Factorial, Response Surface, Mixture and Taguchi designs, where as in Control Charts it monitors processes over time and evaluate their stability [3]. The application of Minitab to Reliability and Survival results in determining a product's life-time characteristics. It uses a wide range of tools that include Distribution Analysis and Accelerated Life Testing.

METHODOLOGY

RESEARCH DATA

The data was collected by Afro Foods Company for the year 2010, from which Minitab was used to analyse it. Based on this data, the analysis was made and is shown in the next section.

DATA ANALYSIS USING COMMAND EXPLAINED MINITAB

The data are put in the first column in the Minitab working interface.

Appendix A

MTB > tsplot c1 this command enables time series plot of the data in column 1, to verify whether or not the data is stationary. The time series plot is shown in Figure-1.

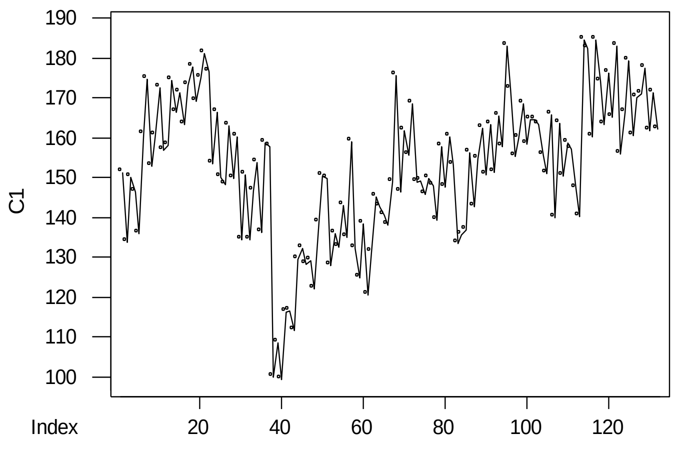

Fig-1: The time series plot

The time series plot shows that the data is not stationary, and has seasonality. Therefore, the data need to be differenced. This can also be evidenced by the trend analysis in Figur-2; the blue line is not horizontal. We therefore difference the data in column 1. Differencing data enables it to be stationary, or this exercise stationarises the data.

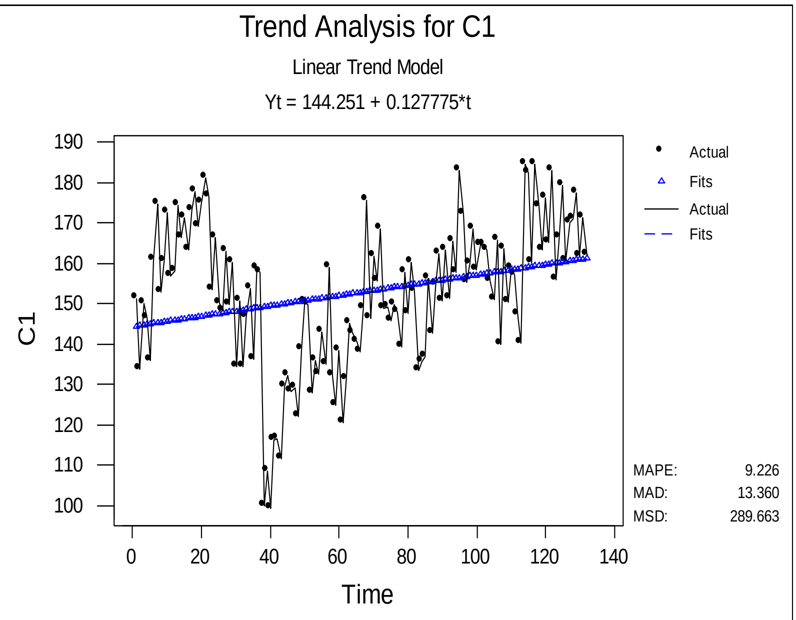

Fig-2: Trend Analysis for C1

MTB > Difference 12 c1 c2 this command differences, at lag 12, the data in column 1 and puts it in column 2. The purpose of differencing the data is to stationarise the data. [That is, to make the data stationary].

MTB > tsplot c2 This command produces a time series plot of the data in c2. The time series plot of the data in c2 is shown in Figure-3.

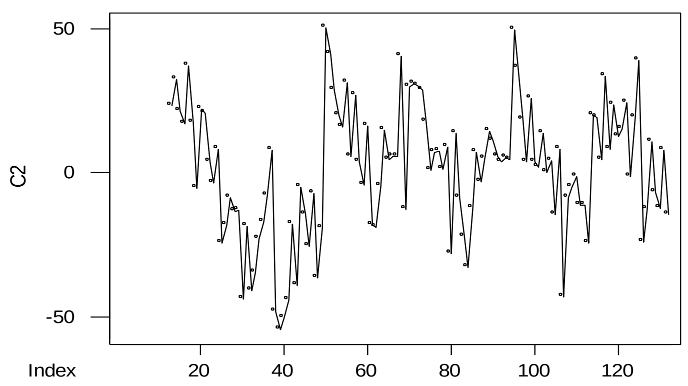

Fig-3

The data is still not stationary; this is evidenced by the blue line which is not horizontal in Figure 4 below. Therefore, the data need to be differenced again, until it becomes stationary.

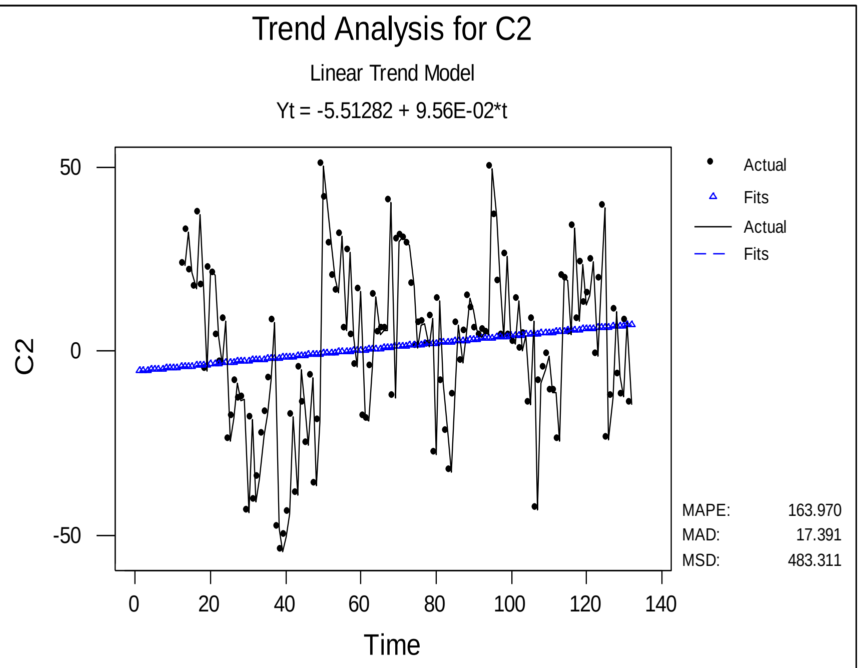

Fig-4: Trend Analysis for C2

MTB > Difference 12 c2 c3 this command enables differencing, at lag 12, the data in c2 and puts it in c3. MTB > tsplot c3 This command enables the time series plot of the data in c3. The data is still not stationary, as shown in Figure 5. We therefore difference again the data in c3.

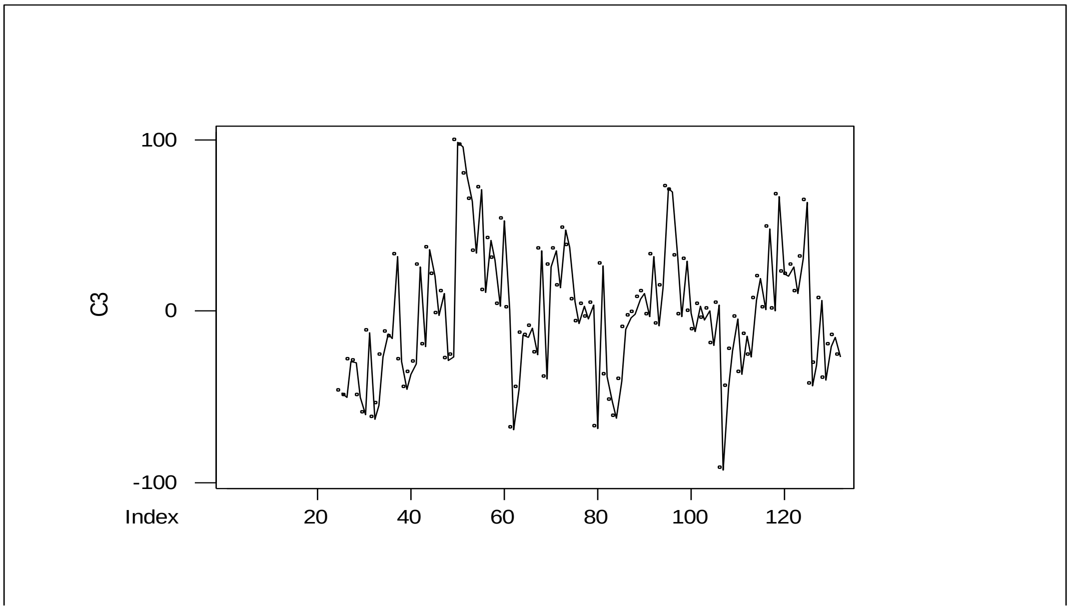

Fig-5

MTB > Difference 12 c3 c4 This command enables differencing, at lag 12, the data in c3 and puts it in c4.

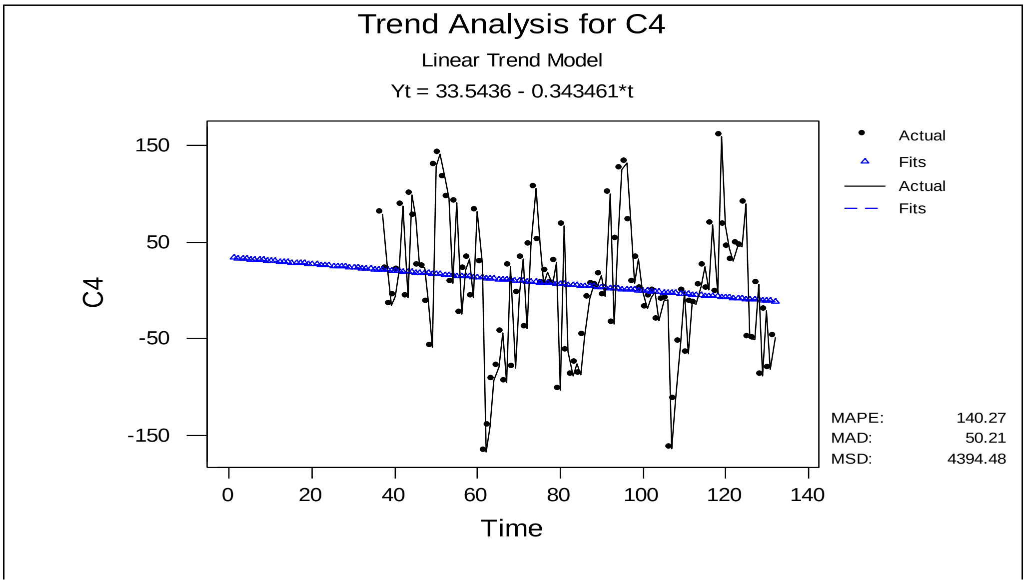

Fig-6: Trend Analysis for C4

The trend analysis in Figure 6 above, of the differenced data in c4 shows that the data is still not stationary. The blue line is not horizontal and not at zero. Therefore we difference again. MTB > Difference 12 c4 c5 This command enables differencing, at lag 12, the data in c4 and puts it in c5.

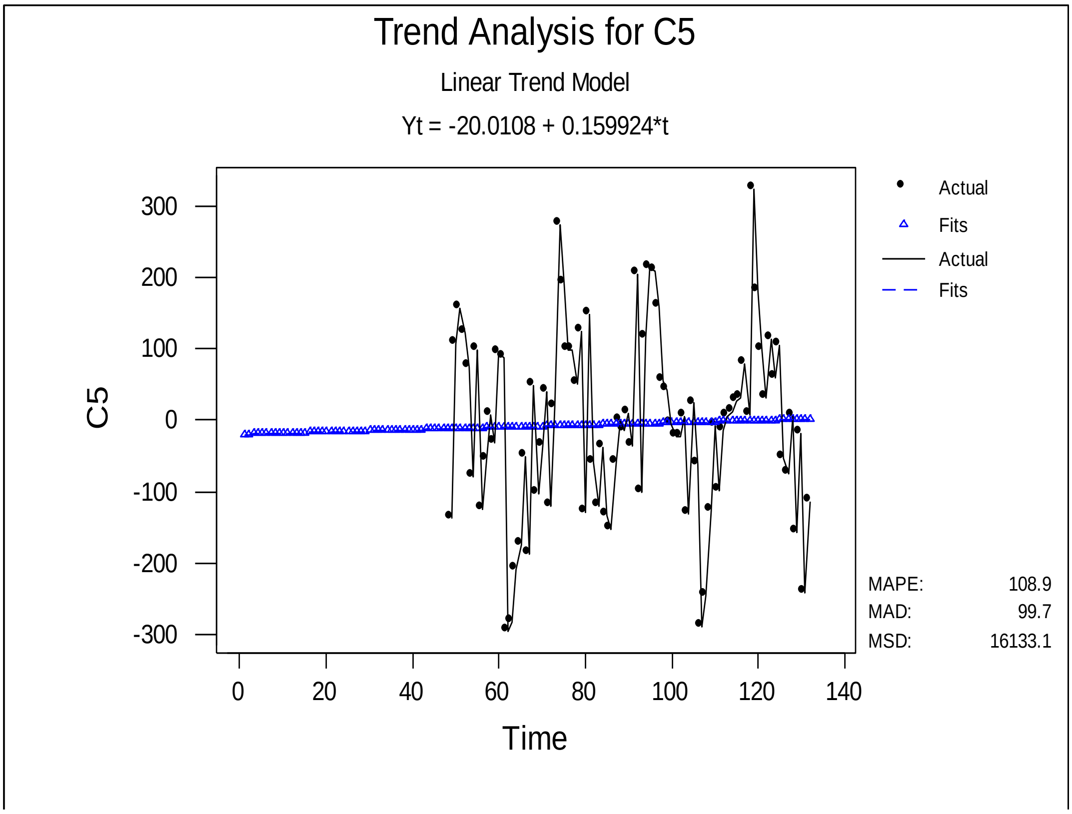

Fig-7: Trend Analysis for C5

The data is almost stationary, but the blue line is tilted, as shown in Figure 7 above. We need to difference again until it becomes stationary. MTB > Difference 12 c5 c6 This command enables differencing, at lag 12, the data in c5 and puts it in c6.

Figure-8 below shows that the data is now stationary, since the data now has no trend. This is because the blue line is horizontal and it is at zero. We now find the Auto-Correlation Function [ACF] and the Partial Auto-Correlation Function [PACF], in order to identify the most likely process the data is coming from.

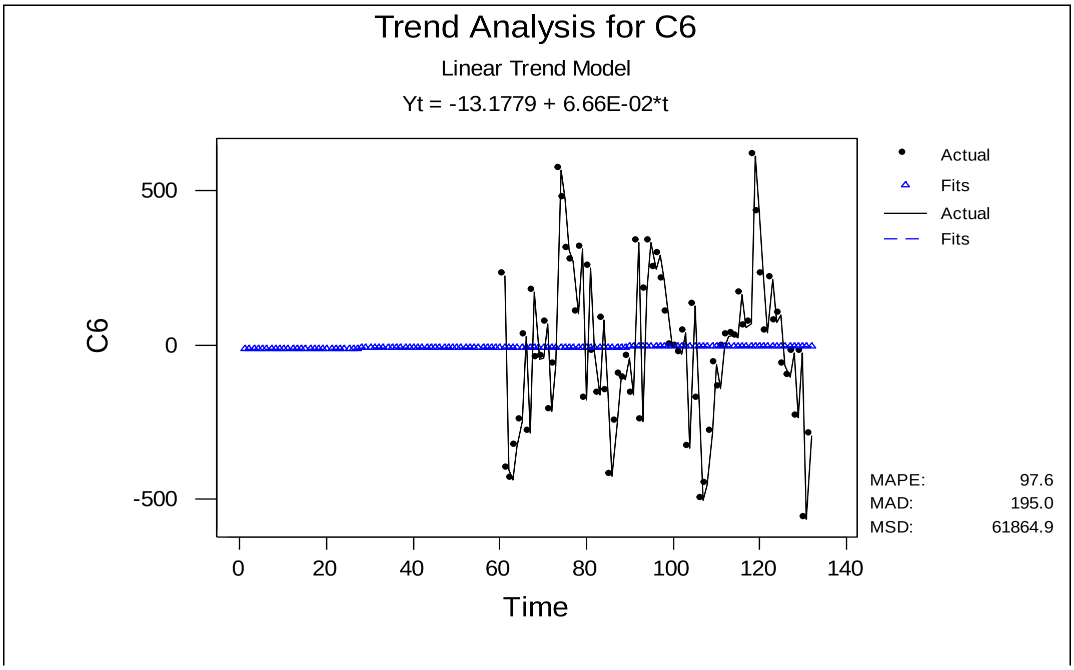

Fig-8: Trend Analysis for C6

MTB > acf c6 This command plots the correlogram of the data in c6. The ACF helps us identify the model, that is, the likely process the data is coming from.

Auto-Correlation Function

ACF of C6

```
         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
           +----+----+----+----+----+----+----+----+----+----+
 1   0.453                       XXXXXXXXXXXX
 2   0.363                       XXXXXXXXXX
 3   0.276                       XXXXXXXX
 4   0.137                       XXXX
 5   0.086                       XXX
 6  -0.075                     XXX
 7  -0.081                     XXX
 8  -0.206                   XXXXXX
 9  -0.332                 XXXXXXXXX
10  -0.379                 XXXXXXXXXX
11  -0.456               XXXXXXXXXXXX
12  -0.737        XXXXXXXXXXXXXXXXXXX
```

Comment on the Auto-Correlation Function [ACF]

The ACF decays slowly from lag 1 to lag 7 and starts to increase from lag 8 to lag 12. This suggests that the model to the data contains an auto-regressive component.

Partial Auto-Correlation Function [PACF]

PACF of C6   This command plots the PACF of the data in c6.

```
         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
           +----+----+----+----+----+----+----+----+----+----+
 1   0.453                       XXXXXXXXXXXX
 2   0.198                       XXXXXX
 3   0.071                       XXX
 4  -0.071                     XXX
 5  -0.021                      XX
 6  -0.167                   XXXXX
 7  -0.021                      XX
```

```
  8  -0.156                XXXXX
  9  -0.210               XXXXXX
 10  -0.175                XXXXX
 11  -0.198               XXXXXX
 12  -0.620       XXXXX
```

Comment on the PACF

The PACF plot shows that there is a significant spike at the first lag. This means that we can fit a tentative model of AR[1]. However this model has to be further investigated to see if it is adequate.

```
MTB > arima 1 5 0 0 0 0 12 c1 c7 c8
Final Estimates of Parameters
Type      Estimate     St. Dev.  t-ratio
AR   1     -0.8152      0.0522    -15.60
Differencing: 5 regular differences
No. of obs.:  Original series 132, after differencing 127
Residuals:    SS =  1121023  [backforecasts excluded]
              MS =     8897  DF = 126
```

Comment

The model seems not to be a good one since it has an MS value of 8897 which is relatively high. So we try another model with a Moving Average [MA] component.

```
MTB > arima 1 5 1 0 0 0 12 c1 c7 c8
Final Estimates of Parameters
Type      Estimate     St. Dev.  t-ratio
AR   1     -0.7877      0.0565    -13.94
MA   1      0.9946      0.0224     44.48
Differencing: 5 regular differences
No. of obs.:  Original series 132, after differencing 127
Residuals:    SS =  369328  [backforecasts excluded]
              MS =    2955  DF = 12
```

Comment

The model seems not to be a good one since it has an MS value of 2955 which is relatively high. So we try another model with a Moving Average [MA] component.

```
MTB > arima 2 5 1 0 0 0 12 c1 c7 c8
Final Estimates of Parameters
Type      Estimate     St. Dev.  t-ratio
AR   1     -1.2520      0.0730    -17.14
AR   2     -0.2587      0.0730     -3.54
MA   1     -0.9938      0.0000 -21819.41

Differencing: 5 regular differences
No. of obs.:  Original series 132, after differencing 127
Residuals:    SS =  2116677  [backforecasts excluded]
              MS =    17070  DF = 124
```

Comment

The model is not good since its MS value of 17070, which is much higher than the previous. So we reject it and try another one.

```
MTB > arima 1 5 2 0 0 0 12 c1 c7 c8
Final Estimates of Parameters
Type      Estimate     St. Dev.  t-ratio
AR   1     -0.7490      0.0623    -12.03
MA   1      1.3657      0.0005   2602.00
MA   2     -0.3798      0.0680     -5.59

Differencing: 5 regular differences
No. of obs.:  Original series 132, after differencing 127
```

Residuals: SS = 246049 [backforecasts excluded]
MS = 1984 DF = 124

Comment

The model seems to be good since its MS value of 1984 relatively much lower than the previous. Since it may not be the best, we try another one.

MTB > arima 2 5 2 0 0 0 12 c1 c7 c8
Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	-1.2926	0.0731	-17.67
AR 2	-0.6914	0.0695	-9.95
MA 1	1.1376	0.0022	524.63
MA 2	-0.1492	0.0902	-1.65

Differencing: 5 regular differences
No. of obs.: Original series 132, after differencing 127
Residuals: SS = 167263 [backforecasts excluded]
MS = 1360 DF = 123

Comment

The model qualifies to be a candidate for forecasting because of its low MS value of 1360. But since it may not be the best one, we try other ones.

MTB > arima 3 5 0 0 0 0 12 c1 c7 c8
Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	-1.8850	0.0708	-26.61
AR 2	-1.6366	0.1102	-14.86
AR 3	-0.6300	0.0704	-8.95

Differencing: 5 regular differences
No. of obs.: Original series 132, after differencing 127
Residuals: SS = 314907 [backforecasts excluded]
MS = 2540 DF = 124

Comment

The model seems not to be good since its MS value of 2540 is relatively much higher than the previous value of 1360. Therefore we reject it and try another one.

MTB > arima 3 5 1 0 0 0 12 c1 c7 c8
Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	-0.1897	0.6384	-0.30
AR 2	0.0206	0.5838	0.04
AR 3	0.2339	0.3318	0.71
MA 1	0.9243	0.6970	1.33

Differencing: 5 regular differences
No. of obs.: Original series 132, after differencing 127
Residuals: SS = 758755 [backforecasts excluded]
MS = 6169 DF = 123

Comment

This model is not good since its MS value of 6169 is higher than the previous. So we reject it and try another one.

MTB > arima 3 5 2 0 0 0 12 c1 c7 c8
Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	-2.0697	0.1007	-20.55
AR 2	-1.7766	0.1506	-11.80
AR 3	-0.6341	0.0830	-7.64

MA 1 0.3599 0.0911 3.95
MA 2 0.6276 0.0681 9.22
Differencing: 5 regular differences
No. of obs.: Original series 132, after differencing 127
Residuals: SS = 177778 [backforecasts excluded]
MS = 1457 DF = 122

Comment

This model is relatively good since its MS value of 1457 is slightly higher than 1360 we found above.

MTB > arima 3 5 3 0 0 0 12 c1 c7 c8
Unable to reduce sum of squares any further
* ERROR * Model cannot be estimated with these data

Comment

This model is not good because it does not give us more information about the parameter values of the model as well as the MS value which, among other parameters determine the candidature of a model. Therefore, we accept ARIMA 2 5 2 0 0 0 12 c1 c7 c8, with MS=1360 and the parameters $\phi 1$ =-1.2926, $\phi 2$ = -0.6914 and $\theta 1$ =1.1376 $\theta 2$ -0.1492. Also, we may take the one below which includes a constant. Although the MS value has increased, it remains the least among all the considered ones. The parameter values also change. Using the table below with parameters $\phi 1$ =-1.3058, $\phi 2$ = -0.6963 and $\theta 1$ =0.9854, $\theta 2$ 0.0315, and after manual calculations, the forecasting model becomes:

$$Y_t = 3.6942\ Y_{t-1} - 4.1673\ Y_{t-2} + 0.4235\ Y_{t-3} + 1.095\ Y_{t-4} + 1.434\ Y_{t-5} - 2.1757\ Y_{t-6} + 0.6963\ Y_{t-7}$$
$$+ a_t - 0.9854\ a_{t-1} - 0.0315\ a_{t-2} - 0.14764$$

MTB > arima 2 5 2 0 0 0 12 c1 c7 c8;
SUBC> constant.
Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	-1.3058	0.0682	-19.14
AR 2	-0.6963	0.0664	-10.49
MA 1	0.9854	0.0002	4184.87
MA 2	0.0315	0.0548	0.57
Constant	-0.14764	0.08303	-1.78

Differencing: 5 regular differences
No. of obs.: Original series 132, after differencing 127
Residuals: SS = 191571 [backforecasts excluded]
MS = 1570 DF = 122

Appendix B

MTB > let c16=loge[c1]
MTB > acf c16

Auto-Correlation Function

ACF of C16

```
        -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
          +----+----+----+----+----+----+----+----+----+----+
  1   0.670                       XXXXXXXXXXXXXXXXXX
  2   0.631                       XXXXXXXXXXXXXXXXX
  3   0.599                       XXXXXXXXXXXXXXXX
  4   0.510                       XXXXXXXXXXXXXX
  5   0.482                       XXXXXXXXXXXXX
  6   0.433                       XXXXXXXXXXXX
  7   0.434                       XXXXXXXXXXXX
  8   0.384                       XXXXXXXXXXX
  9   0.318                       XXXXXXXXX
 10   0.286                       XXXXXXXX
```

```
 11   0.219                XXXXXX
 12   0.208                XXXXXX
```

Comment on the ACF

The ACF decays exponentially, though some insignificant spikes exist at the bottom. Now we difference it once, at lag 12 and plot the ACF again.

MTB > diff 12 c16 c17

MTB > acf c17

Autocorrelation Function

ACF of C17

```
        -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
          +----+----+----+----+----+----+----+----+----+----+
  1   0.557                    XXXXXXXXXXXXXXX
  2   0.460                    XXXXXXXXXXXX
  3   0.421                    XXXXXXXXXXXX
  4   0.300                    XXXXXXXXX
  5   0.269                    XXXXXXXX
  6   0.211                    XXXXXX
  7   0.157                    XXXXX
  8   0.116                    XXXX
  9  -0.025                   XX
 10  -0.086                   XXX
 11  -0.195                 XXXXXX
 12  -0.356              XXXXX
```

Comment on the ACF

The ACF decays exponentially, with insignificant spikes at the bottom. Now we plot the PACF.

MTB > pacf c17

Partial Auto-Correlation Function

PACF of C17

```
        -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
          +----+----+----+----+----+----+----+----+----+----+
  1   0.557                    XXXXXXXXXXXXXXX
  2   0.217                    XXXXXX
  3   0.152                    XXXXX
  4  -0.037                   XX
  5   0.038                    XX
  6  -0.014                    X
  7  -0.016                    X
  8  -0.026                   XX
  9  -0.179                 XXXXX
 10  -0.098                   XXX
 11  -0.172                 XXXXX
 12  -0.263               XXXXXXXX
```

Comment on the PACF

The PACF has an insignificant spike at the first lag suggesting that we can fit an ARIMA model with an AR component present.

MTB > arima 1 0 0 0 1 1 12 c16 c7 c8

Final Estimates of Parameters

```
Type      Estimate   St. Dev.  t-ratio
AR   1      0.4929     0.0720     6.85
SMA 12      0.5020     0.0981     5.12
```

Differencing: 0 regular, 1 seasonal of order 12

No. of obs.: Original series 132, after differencing 120

Residuals: SS = 1.35366 [backforecasts excluded]

MS = 0.01147 DF = 118

Comment on the Table

The model seems okay, with ϕ_1 =0.4929, t-ratio= 6.85 and MS=0.01147. We now try to fit another model which may reduce the value of MS.

MTB > arima 1 0 1 0 1 1 12 c16 c7 c8
Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	0.9335	0.0428	21.83
MA 1	0.5384	0.1000	5.38
SMA 12	0.8776	0.0748	11.73

Differencing: 0 regular, 1 seasonal of order 12
No. of obs.: Original series 132, after differencing 120
Residuals: SS = 0.902030 [backforecasts excluded]
MS = 0.007710 DF = 117

Comment on the Table

The model seems okay, with ϕ_1 =0.9335, t-ratio= 21.83, θ_1 =0.5284, t-ratio= 5.38 and MS=0.007710, which is lower than the previous value. To see that it is a good model, we now test the following: normality, independence, randomness and constant variance of the residuals.

Test of constant Variance

MTB > dotplot c7
Character Dotplot
12 Points missing or out of range

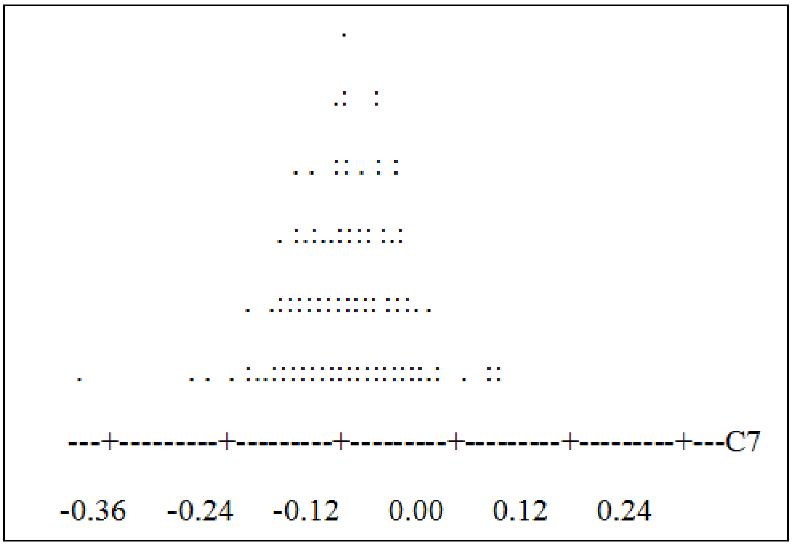

Fig-9

Comment

Since the plot in Figure-9 is funnel-shaped, then the principle of heteroskedasticity is not violated. This means that there is no constant variance.

Test of Randomness

MTB > nscores c8 c15
MTB > plot c8*c15

- NOTE * N missing = 12

The command gives a normal scores plot of the fits in c8 versus the squares of the residuals in c15.

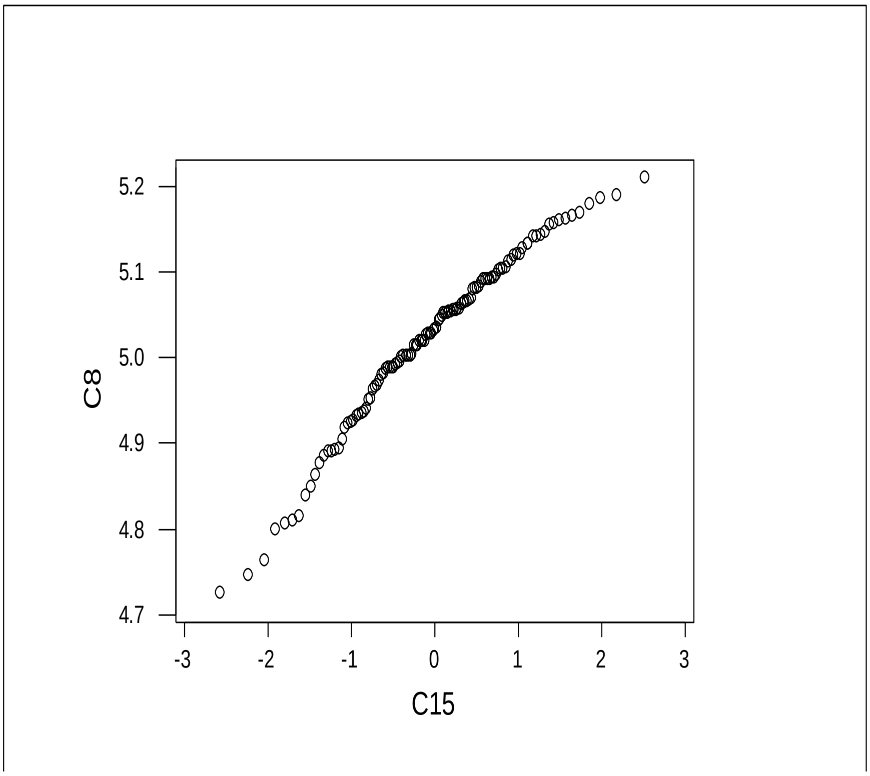

Fig-9:

Comment

The line in **Figure-10** is almost straight, verifying that the principle of randomness of residuals is not violated. That is, the residuals are randomly distributed.

Test for Normality

MTB > hist c7

- NOTE * N missing = 12

This command plots a histogram of the residuals in c7.

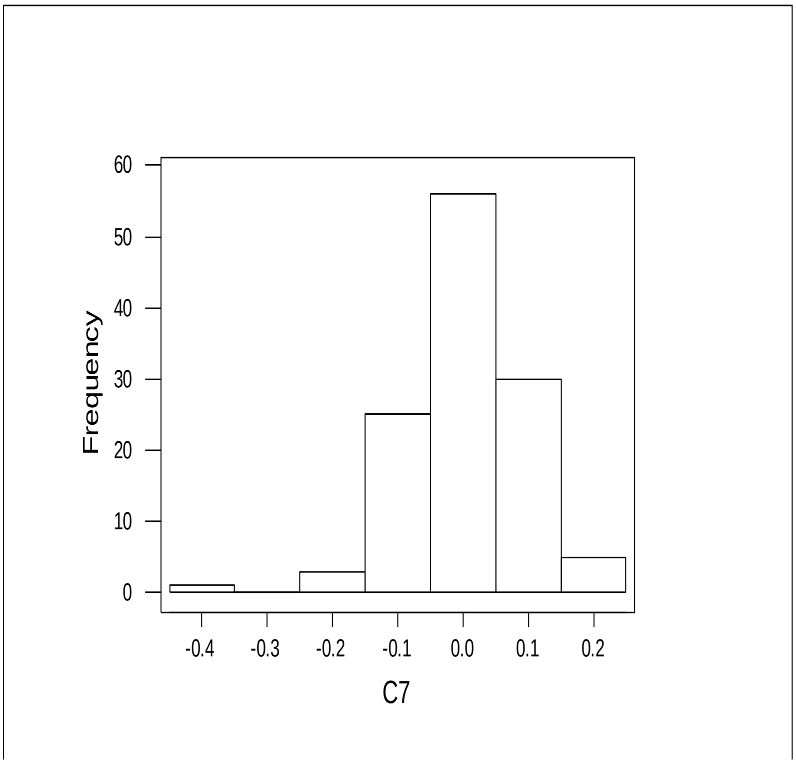

Fig-10

MTB > %NormPlot C8.

This command gives a normal probability plot of fits in c8, which is in Figure-12.

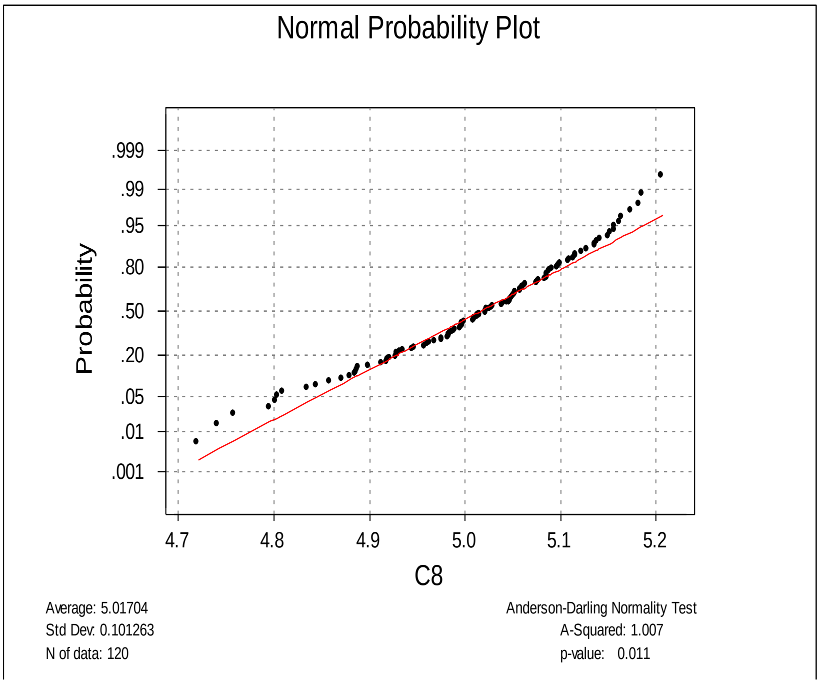

Fig-11: Normal Probability Plot

Comment

The plots in Figures 11, 12, the histogram and the normal probability plot, respectively of the residuals show that the residuals are normally distributed.

MTB > lag c7 c20
MTB > plot c7*c20

- NOTE * N missing = 13

The command above gives a scatter plot of the residuals in c7 versus time.

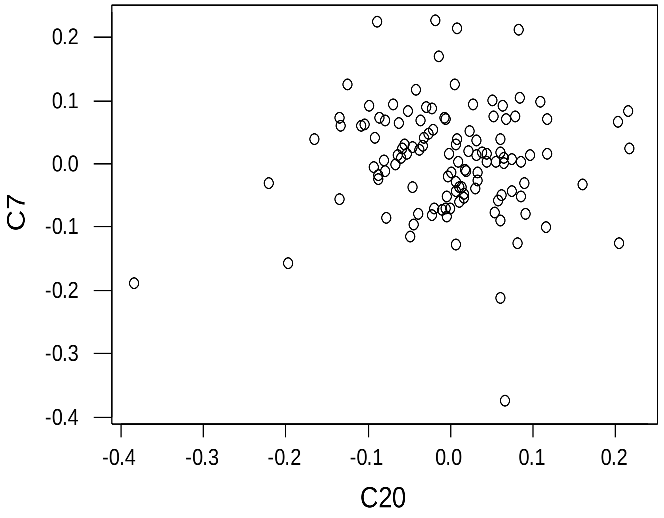

Fig-12

Comment

Figure 13 shows that there is no pattern that is followed by the residuals, hence no violation of randomness of residuals.

MTB > regr c15 1 c8
Regression Analysis
The regression equation is
C15 = 5.71 - 1.14 C8
120 cases used 12 cases contain missing values

Predictor	Coef	Stdev	t-ratio	p
Constant	5.706	4.486	1.27	0.206
C8	-1.1373	0.8940	-1.27	0.206

s = 0.9876 R-sq = 1.4% R-sq [adj] = 0.5%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	1.5784	1.5784	1.62	0.206
Error	118	115.0819	0.9753		
Total	119	116.6603			

MTB > regress c7 1 c20

This command compiles the tables that analyses the residuals in column 7. That is, it focuses on the regression analysis of the residuals stored in column 7.

Regression Analysis
The regression equation is
C7 = 0.00198 - 0.0007 C20
119 cases used 13 cases contain missing values

Predictor	Coef	Stdev	t-ratio	p
Constant	0.001978	0.007988	0.25	0.805
C20	-0.00066	0.09190	-0.01	0.994

s = 0.08707 R-sq = 0.0% R-sq[adj] = 0.0%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	0.000000	0.000000	0.00	0.994
Error	117	0.886906	0.007580		
Total	118	0.886906			

MTB > runs 0 c7

Runs Test

C7

K = 0.0000

The observed no. of runs = 70

The expected no. of runs = 60.5833

65 Observations above K 55 below

The test is significant at 0.0824

Cannot reject at alpha = 0.05

Report for Afro Foods Company

INTRODUCTION

The report focuses on the likely sales of rice in Zimbabwe for the first three months of 2011 as requested by the Afro Foods management. The report is based on ARIMA model and the Explanatory model that were built using the previous data of the sales of rice.

Board

The models for forecasting the likely sales of rice in Zimbabwe for the first three months of 2011 were built basing on the past data of rice sales. The ARIMA model that was built basing on this data is found in Appendix A, from which the formula for calculating the approximate number of sales was derived. The formula is given below:

$$Y_t = 3.6942\,Y_{t-1} - 4.1673\,Y_{t-2} + 0.4235\,Y_{t-3} + 1.095\,Y_{t-4} + 1.434\,Y_{t-5} - 2.1757\,Y_{t-6} + 0.6963\,Y_{t-7}$$
$$+ a_t - 0.9854\,a_{t-1} - 0.0315\,a_{t-2} - 0.14764 \quad ,$$

Where, Y_t is the total number of sales in each month. This tells us that, if we want to forecast the sales of January, we base on the sales of rice for the past seven months, that is, from June 2010 to December 2010. For February we take rice sales from July 2010 to January 2011, etc. Several models were tried for relevance to the rice data, but they failed because of their failure to meet the selecting criteria. The one chosen had favourable values of the t-ratios and a lower MS value which guides us how close a model is in fitting the given data.

However, the other formula was constructed from Appendix B, which is based on the Explanatory model. The regression equation is

Y_t = 5.71 - 1.14 X_t ,

where Y_t is defined as before and X_t are the fits. This formula is based on the logarithms of the data of the past sales.

CONCLUSION AND RECOMMENDATIONS

Considering the time period, three months is a small period of time, as such we therefore recommend the use of the ARIMA model which is mostly used for short-term forecasting. Explanatory model is often used for long-term forecasting [4].

REFERENCES

1. Bersimis S, Psarakis S, Panaretos J. Multivariate statistical process control charts: an overview. Quality and Reliability engineering international. 2007 Aug 1;23(5):517-43.
2. Callao MP, Rius A. Time series: a complementary technique to control charts for monitoring analytical systems. Chemometrics and intelligent laboratory systems. 2003 Apr 28;66(1):79-87.
3. Hotelling H. Multivariate quality control. Techniques of statistical analysis. 1947.
4. Brian Hwarng H, Wang Y. Shift detection and source identification in multivariate autocorrelated processes. International Journal of Production Research. 2010 Feb 1;48(3):835-59.