FORECASTING TIME SERIES HOMEWORK 1

TEAM G

Vratul Kapur
Paul Jacques-Mignault
Sheena Miles
Ashley O'Mahony
Stavros Tsentemeidis
Karl Westphal
Irune Maury Arrue

SERIES ANALYSIS SUMMARY

		Series 1	Series 2	Series 3	Series 4	Series 5	Series 6	Series 7
Original scale	Stationary	yes	no	no	no	yes	yes	no
	Marginal Normal distribution	no	no	no	no	no	no	no
	WN	yes	no	no	no	no	no	no
	SWN	?	no	no	no	no	no	no
	GWN	no	no	no	no	no	no	no
	Linear model	no	yes	yes	yes	yes	yes	yes
	Nonlinear model	no	?	?	?	?	?	?
	Transformations (d)	d=0	d=1	d=1	d=2	d=0	d=0	d=1
Transformed data	Stationary	-	yes	yes	no	-	-	yes
	Marginal Normal distribution	-	yes	yes	no	-	-	no
	WN	-	yes	no	no	-	-	no
	SWN	-	yes	no	no	-	-	no
	GWN	-	yes	no	no	-	-	no
	Linear model	-	no	yes	yes	-	-	yes
	Nonlinear model	-	no	?	?	-	-	?
	Transformations (d)	-	d=0	d=0	d=1	-	-	d=0
Transformed data	Stationary	-	-	-	yes	-	-	-
	Marginal Normal distribution	-	-	-	yes	-	-	-
	WN	-	-	-	no	-	-	-
	SWN	-	-	-	no	-	-	-
	GWN	-	-	-	no	-	-	-
	Linear model	-	-	-	yes	-	-	-
	Nonlinear model	-	-	-	?	-	-	-
	Transformations (d)	-	-	-	d=0	-	-	-

Table 1: Time Series Analysis Summary

Assumptions for our analysis:

- Comparative statistics are based on a 95% confidence interval.
- In any case we might think the series might not be stationary in the variance, we continue our analysis as our data points contain negative values making it inconvenient to process with the log function.

SERIES 1

Original

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are stationary through the mean and the variance. So, no transformation is needed.
- **Normality:** even though through the histogram we cannot say for sure if the data is normal, we can definitely say that they are not through the Shapiro test (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots we believe that our data is White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
    - As the data is WN but not normal, there cannot be Gaussian White Noise.
    - As the data is originally WN, we can explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one. We see that the squares are also WN, which means that there is no non-linear quadratic model for our data.
    - Regarding Strict White Noise, we cannot make any assumption as there might be a relationship that is nonlinear apart from quadratic.

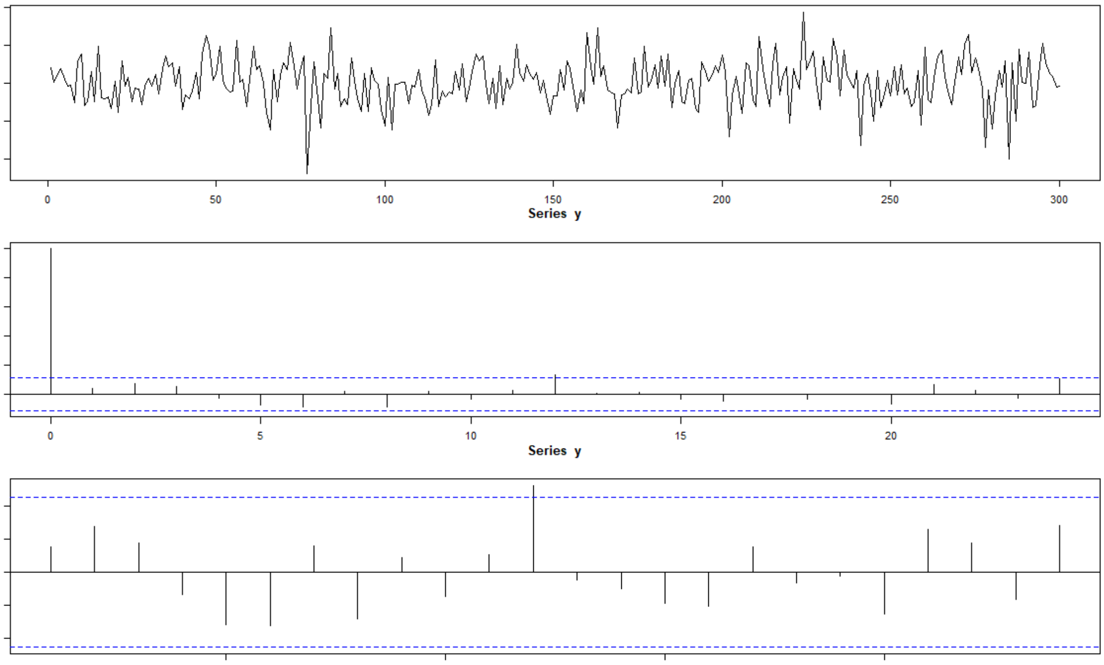


Figure 1: Plot, ACF and PACF of Series 1 Data

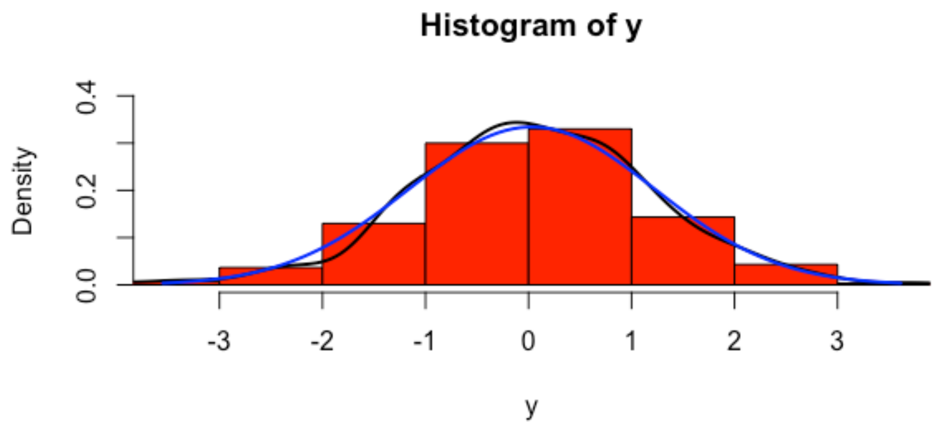


Figure 2: Histogram of Series 1 Data

```
> mean(y)
[1] 0.03205
> sd(y)
[1] 1.193712
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 0
> #Normality
> skewness(y)
[1] -0.2795834
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 4.040454
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.98868, p-value = 0.01941

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 18.804, df = 20, p-value = 0.5346
```

Figure 3: Formal Analysis of Series 1 Data with R

SERIES 2

Original

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are not stationary through the mean. So, one transformation is needed and specifically one difference.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is not normal (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value below 0.05).
    - As the data is not WN, there cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

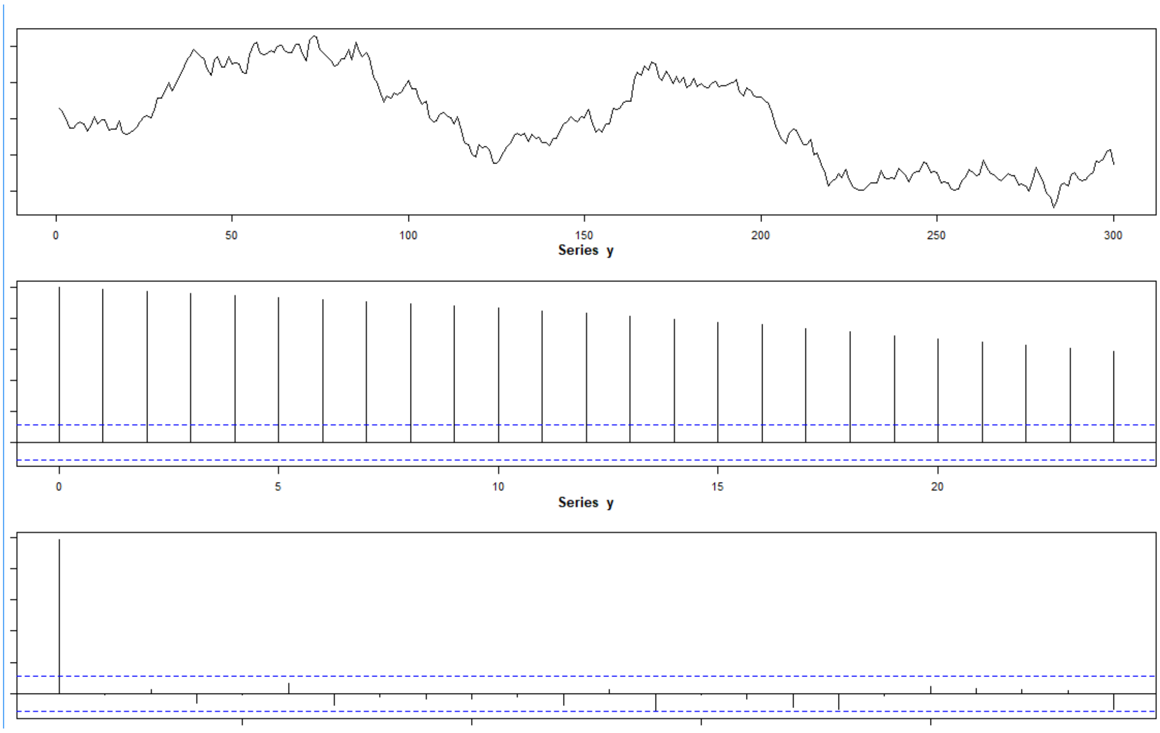


Figure 4: Plot, ACF and PACF of Series 2 Data

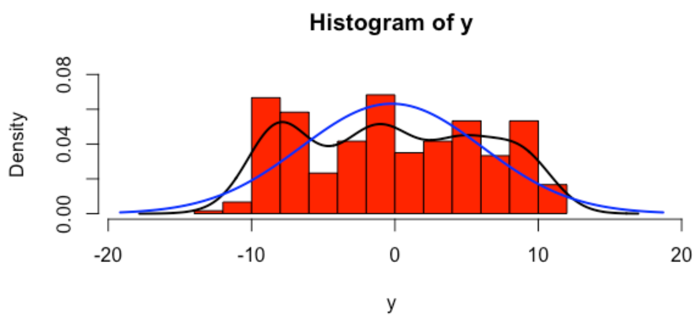


Figure 5: Histogram of Series 2 Data

```
> mean(y)
[1] -0.22853
> sd(y)
[1] 6.31268
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 1
> #Normality
> skewness(y)
[1] 0.04883792
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 1.765163
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.94921, p-value = 1.133e-08

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 4509.5, df = 20, p-value < 2.2e-16
```

Figure 6: Formal Analysis of Series 2 Data with R

Transformed

- **Stationary:** by applying one difference at our original data, we can now observe through both the plots and the ADF test, that our data is finally stationary. So, no further transformation is needed at this point.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is normal (p-value above 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
  - As the data is WN and normal, there can be Gaussian White Noise. So, as they are GWN, they are also SWN.
  - As the data is WN, we can say that we cannot have a linear model for our time series.
  - As the data is originally WN, we can explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one. We see that the squares are also WN, which means that there is no non-linear quadratic model for our data.

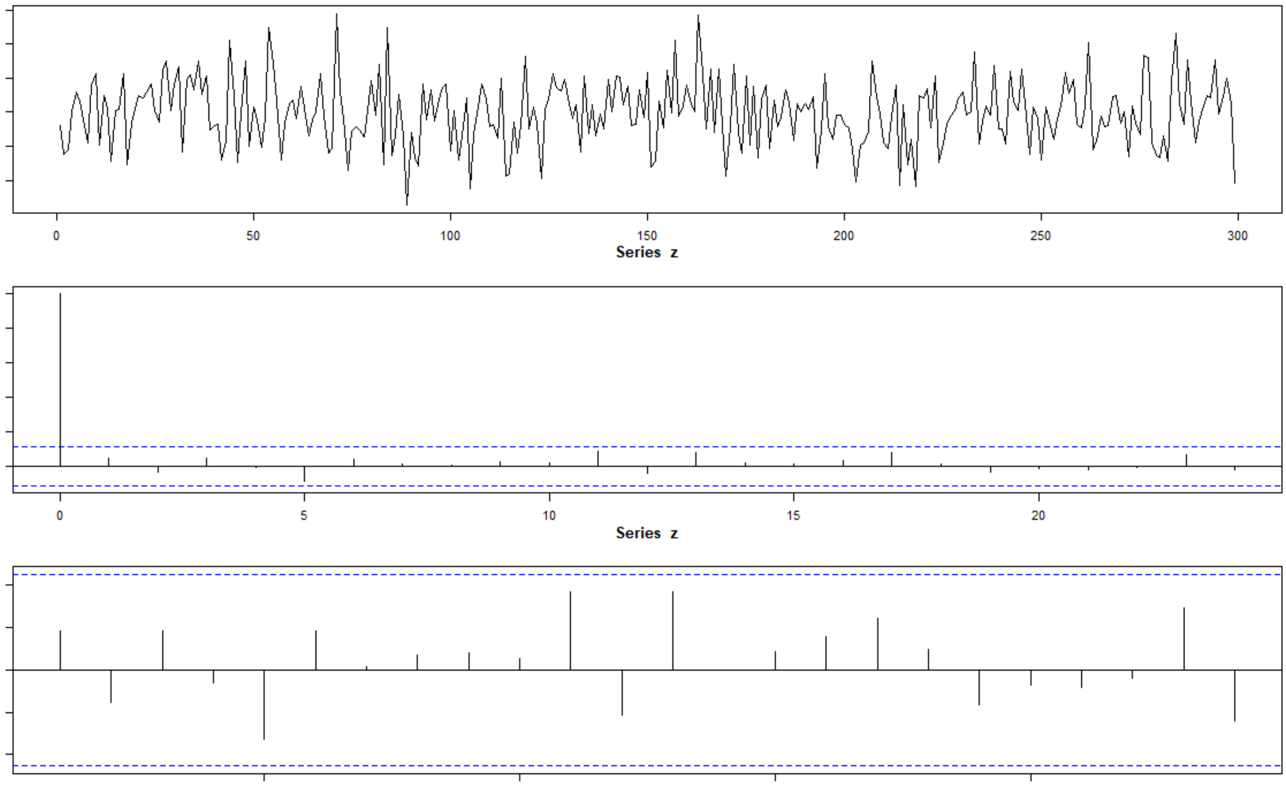


Figure 7: Plot, ACF and PACF of Series 2 Transformed Data

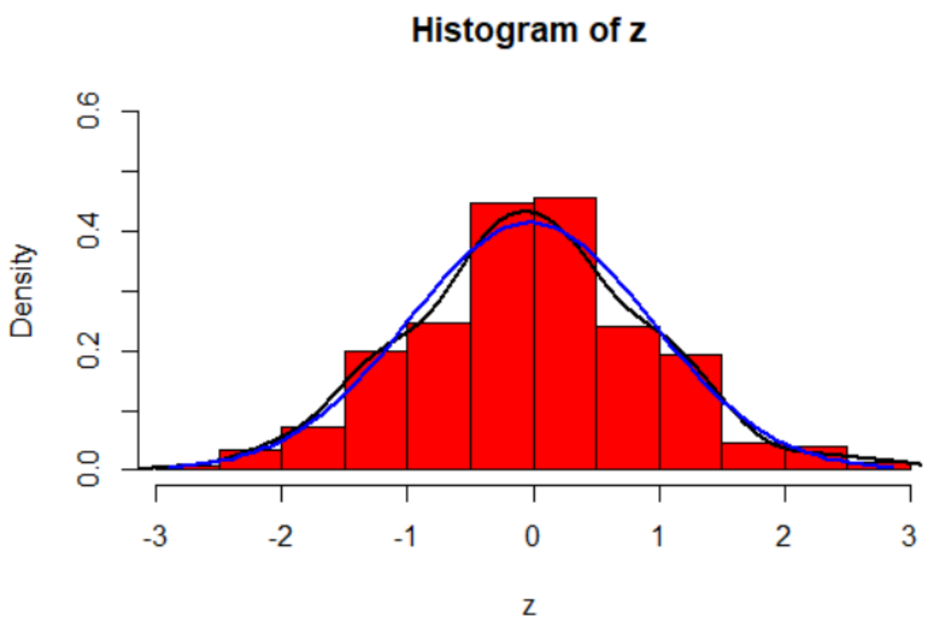


Figure 8: Histogram of Series 2 Transformed Data

```
> mean(z)
[1] -0.02598662
> sd(z)
[1] 0.9620915
> ndiffs(z, alpha=0.05, test=c("adf"))
[1] 0
> Box.test (z, lag = 20, type="Ljung")

        Box-Ljung test

data:  z
X-squared = 12.452, df = 20, p-value = 0.8996

> skewness(y)
[1] 0.04883792
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 1.765163
attr(,"method")
[1] "moment"
> shapiro.test(z)

        Shapiro-Wilk normality test

data:  z
W = 0.99619, p-value = 0.69
```

Figure 9: Formal Analysis of Series 2 Transformed Data with R

SERIES 3

Original

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are not stationary through the mean. So, one transformation is needed and specifically one difference.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is not normal (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value below 0.05). Another factor that led us to this conclusion is the distance of the mean from 0, which is one of the 3 factors that define a White Noise.
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

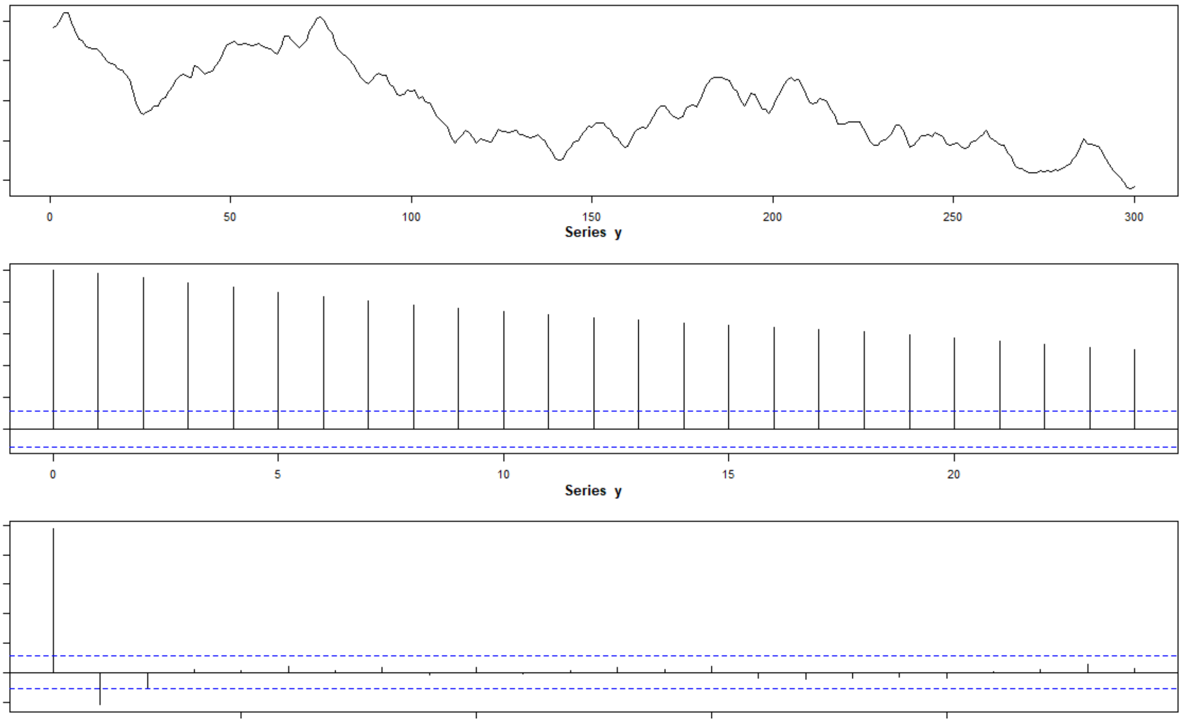


Figure 10: Plot, ACF and PACF of Series 3 Data

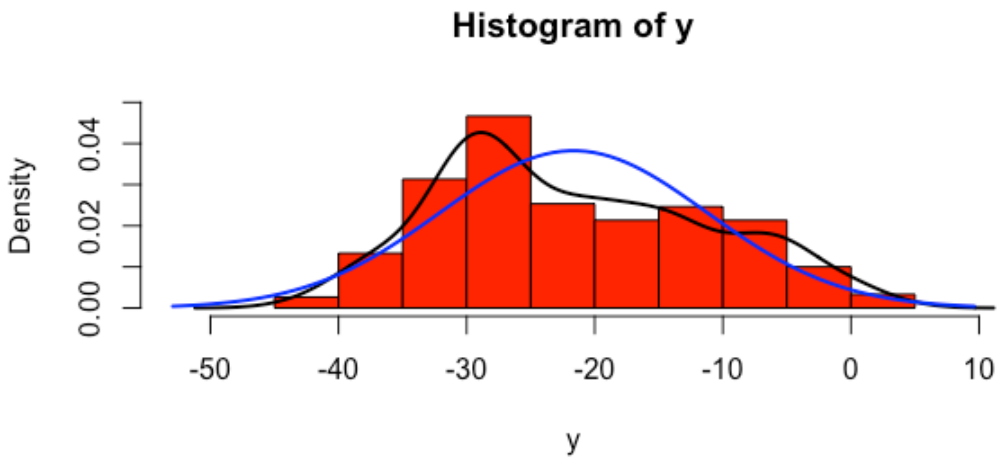


Figure 11: Histogram of Series 3 Data

```
> mean(y)
[1] -21.65212
> sd(y)
[1] 10.42905
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 1
> #Normality
> skewness(y)
[1] 0.3495993
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 2.197106
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.9645, p-value = 1.019e-06

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 3598.9, df = 20, p-value < 2.2e-16
```

Figure 12: Formal Analysis of Series 3 Data with R

Transformed

- **Stationary:** by applying one difference at our original data, we can now observe through both the plots and the ADF test, that our data is finally stationary. So, no further transformation is needed at this point.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is normal (p-value above 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

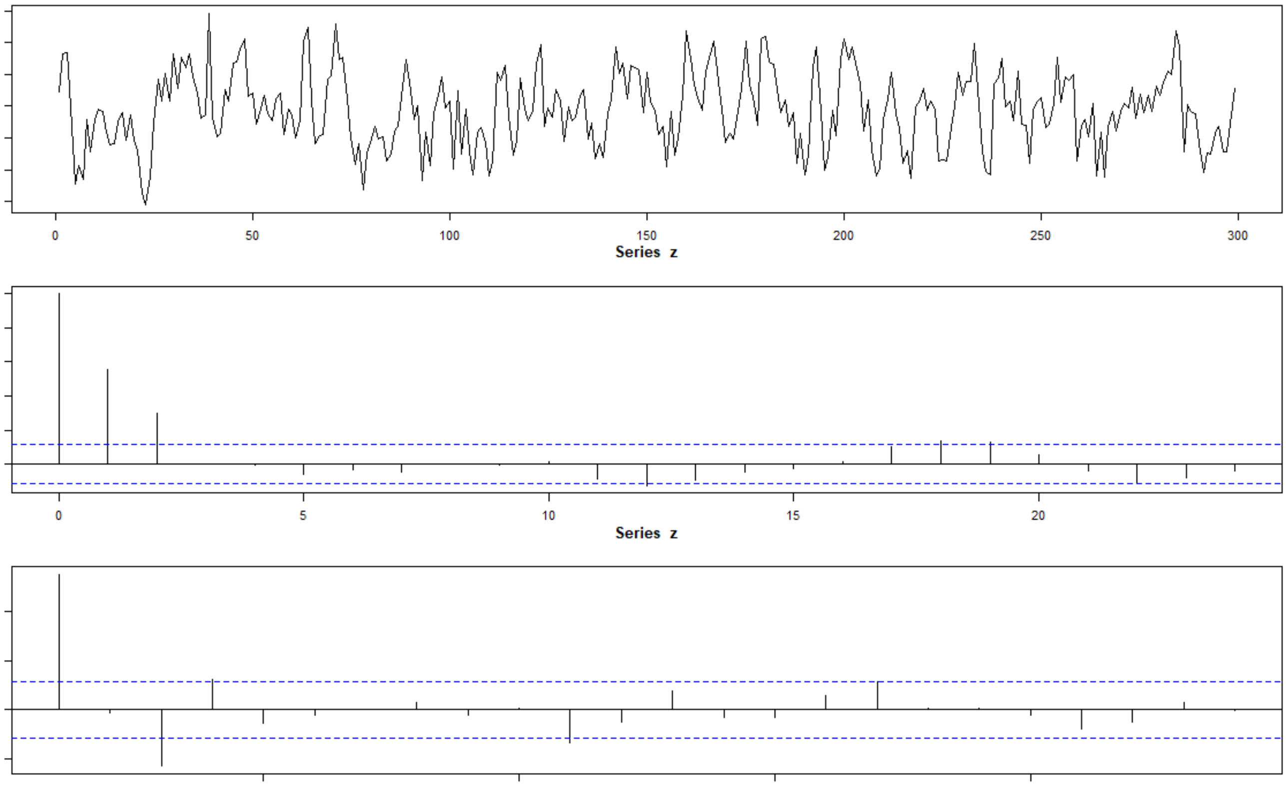


Figure 13: Plot, ACF and PACF of Series 3 Transformed Data

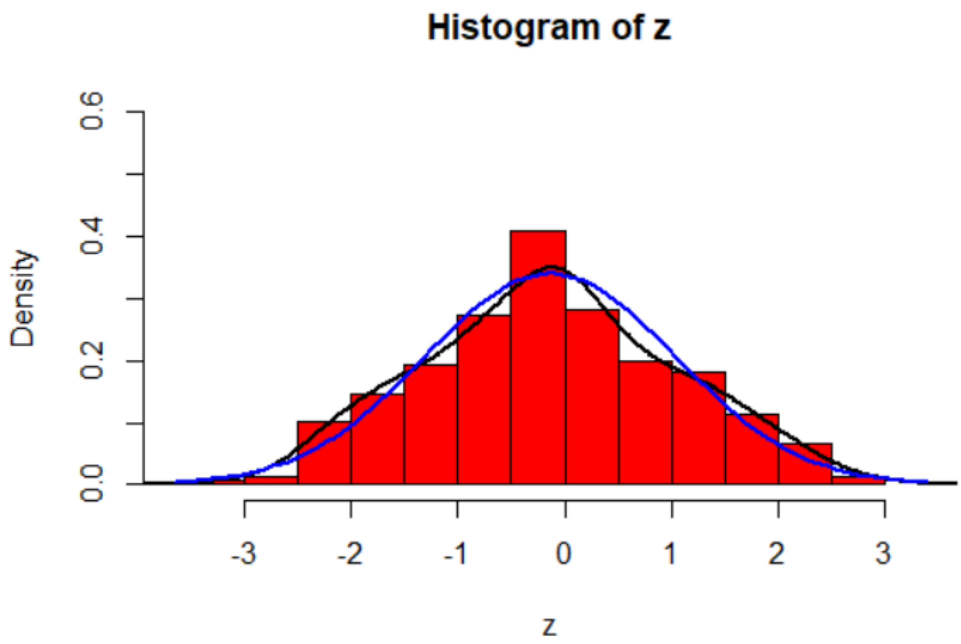


Figure 14: Histogram of Series 3 Transformed Data

```
> mean(z)
[1] -0.1338495
> sd(z)
[1] 1.172368
> ndiffs(z, alpha=0.05, test=c("adf"))
[1] 0
> Box.test (z, lag = 20, type="Ljung")

        Box-Ljung test

data:  z
X-squared = 148.71, df = 20, p-value < 2.2e-16

> skewness(y)
[1] 0.3495993
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 2.197106
attr(,"method")
[1] "moment"
> shapiro.test(z)

        Shapiro-Wilk normality test

data:  z
W = 0.99344, p-value = 0.2178
```

Figure 15: Formal Analysis of Series 3 Transformed Data with R

SERIES 4

Original

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are not stationary through the mean. So, two transformations are needed and specifically two differences.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is not normal (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value below 0.05). Another factor that led us to this conclusion is the distance of the mean from 0, which is one of the 3 factors that define a White Noise.
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

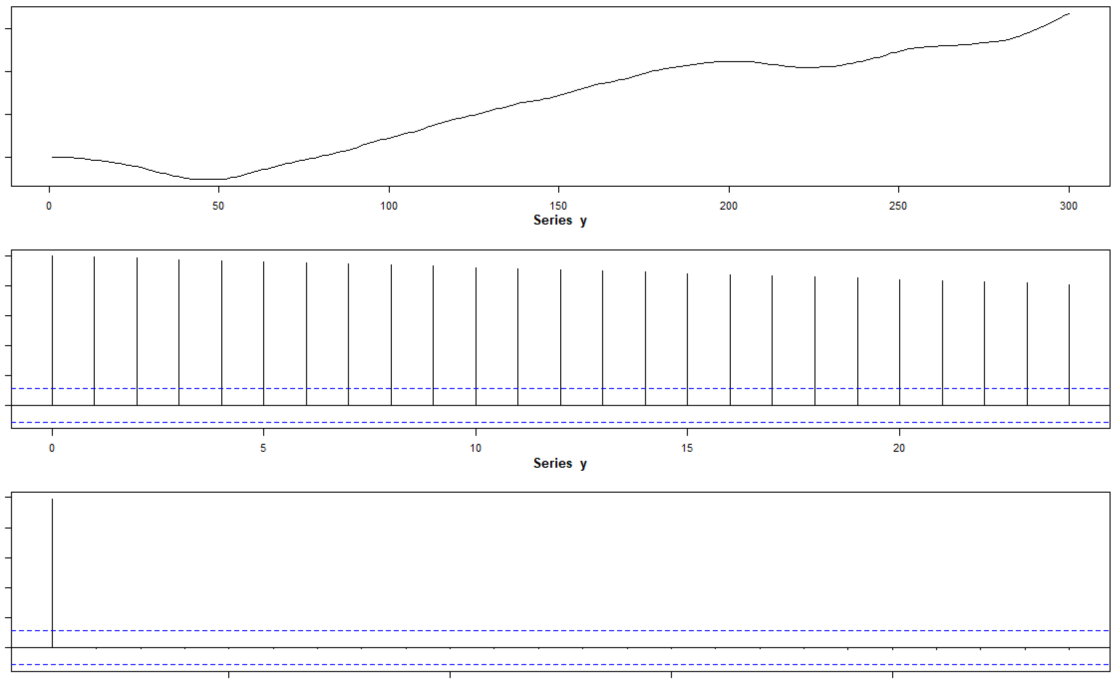


Figure 16: Plot, ACF and PACF of Series 2 Data

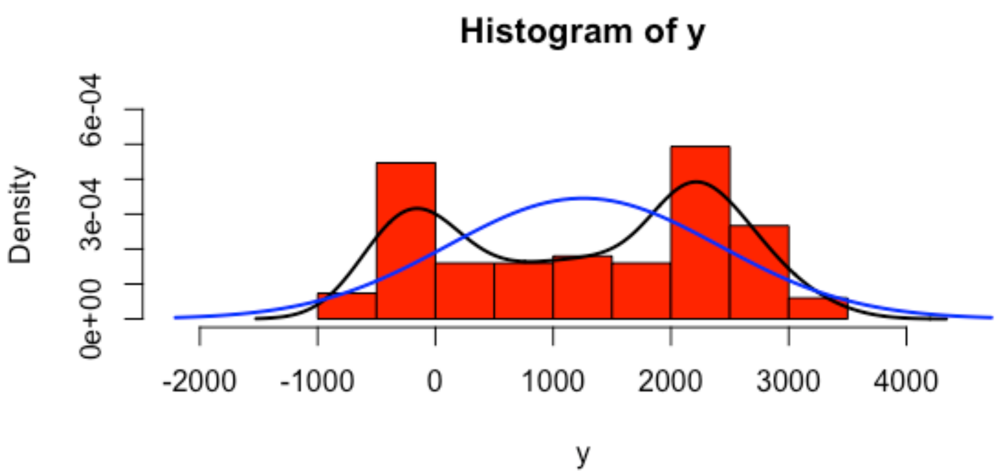


Figure 17: Histogram of Series 4 Data

```
> mean(y)
[1] 1256.843
> sd(y)
[1] 1154.857
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 2
> #Normality
> skewness(y)
[1] -0.1488262
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 1.559977
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.91161, p-value = 2.779e-12

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 5290.9, df = 20, p-value < 2.2e-16
```

Figure 18: Formal Analysis of Series 4 Data with R

Transformed 1st difference

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are not stationary through the mean. So, two transformations are needed and specifically two differences.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is not normal (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value below 0.05). Another factor that led us to this conclusion is the distance of the mean from 0, which is one of the 3 factors that define a White Noise.
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

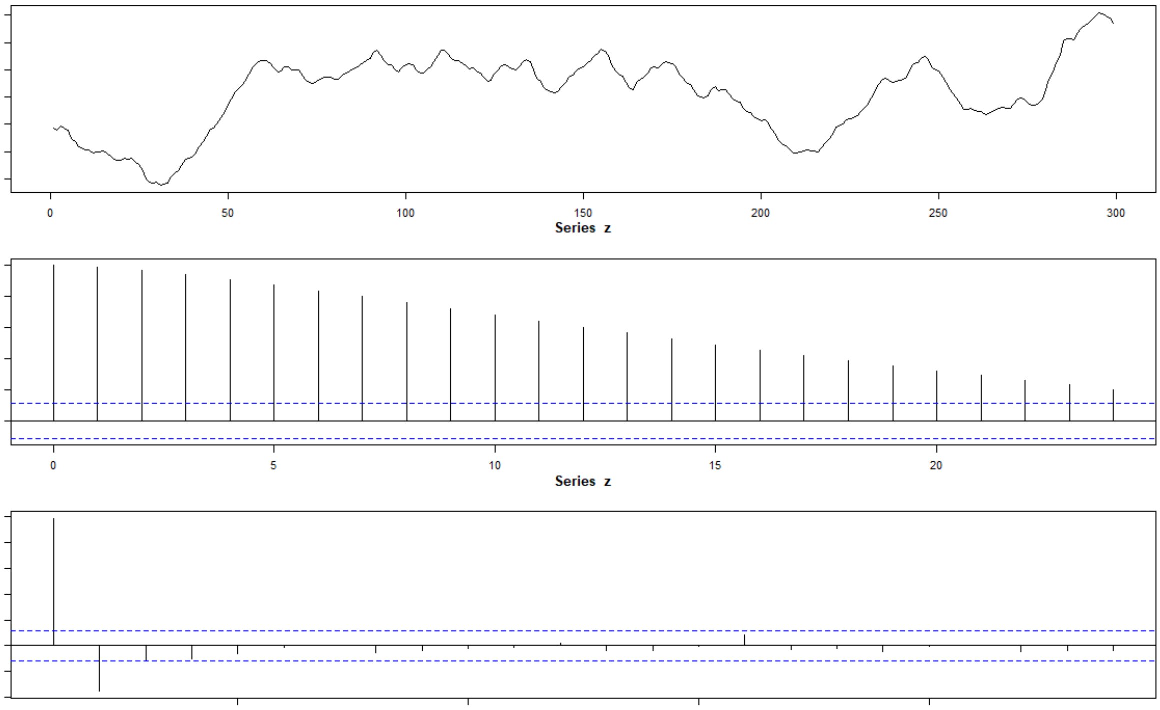


Figure 19: Plot, ACF and PACF of Series 4 First Transformed Data

```
> mean(z)
[1] 11.17006
> sd(z)
[1] 13.82722
> ndiffs(z, alpha=0.05, test=c("adf"))
[1] 1
> Box.test (z, lag = 20, type="Ljung")

        Box-Ljung test

data:  z
X-squared = 2974.9, df = 20, p-value < 2.2e-16

> skewness(y)
[1] -0.1488262
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 1.559977
attr(,"method")
[1] "moment"
> shapiro.test(z)

        Shapiro-Wilk normality test

data:  z
W = 0.94851, p-value = 9.871e-09
```

Figure 20: Formal Analysis of Series 4 First Transformed Data with R

Transformed 2nd difference

- **Stationary:** by applying one difference at our original data, we can now observe through both the plots and the ADF test, that our data is finally stationary. So, no further transformation is needed at this point.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is normal (p-value above 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

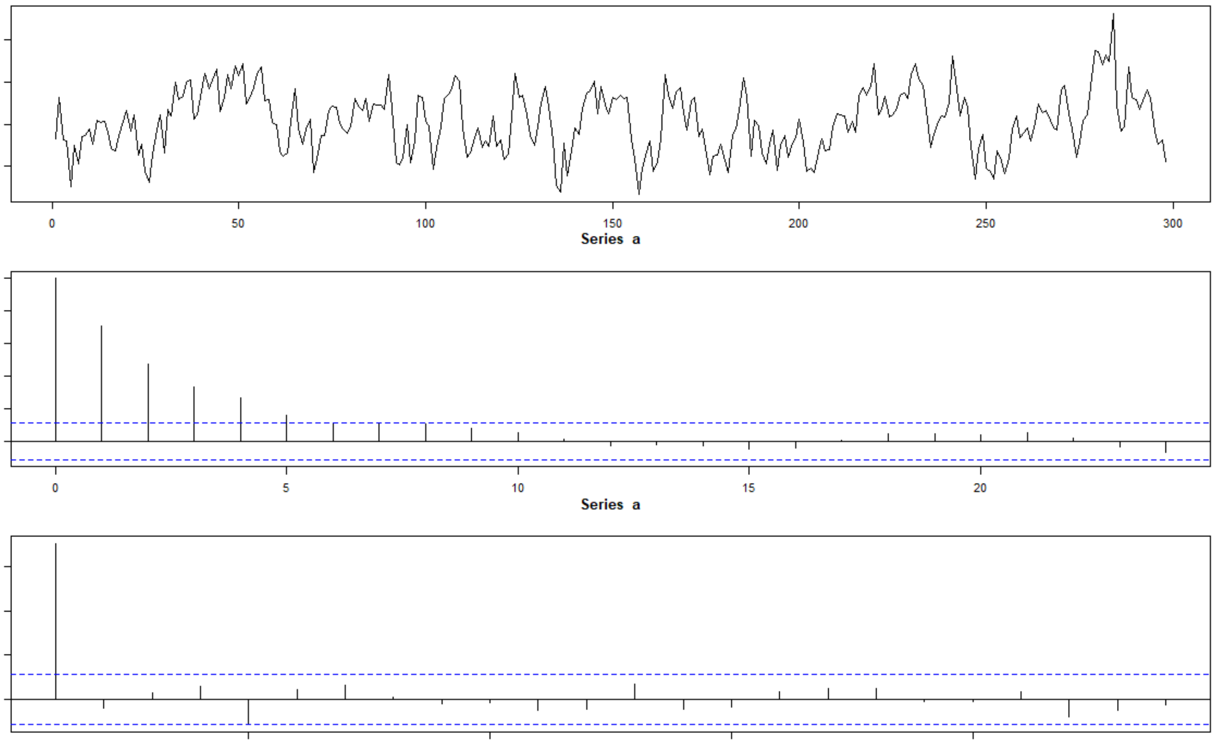


Figure 21: Plot, ACF and PACF of Series 4 Second Transformed Data

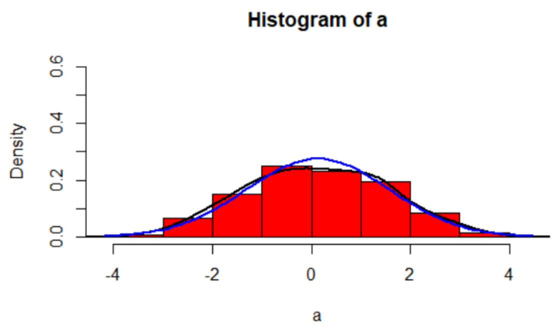


Figure 22: Histogram of Series 4 Second Transformed Data

```
> mean(a)
[1] 0.1287819
> sd(a)
[1] 1.445643
> ndiffs(a, alpha=0.05, test=c("adf"))
[1] 0
> Box.test (a, lag = 20, type="Ljung")

        Box-Ljung test

data:  a
X-squared = 297.76, df = 20, p-value < 2.2e-16

> skewness(a)
[1] 0.1289298
attr(,"method")
[1] "moment"
> kurtosis(a,method=c("moment"))
[1] 2.761813
attr(,"method")
[1] "moment"
> shapiro.test(a)

        Shapiro-Wilk normality test

data:  a
W = 0.99438, p-value = 0.3416
```

Figure 23: Formal Analysis of Series 4 Second Transformed Data with R

SERIES 5

Original

- Stationary: both through the time series plot and the ADF test, we can verify that our data are stationary through the mean and the variance. So, no transformation is needed.
- Normality: even though through the histogram we cannot say for sure if the data is normal, we can definitely say that they are not through the Shapiro test (p-value below 0.05).
- White Noise: through the ACF and PACF plots, we believe that our data is White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

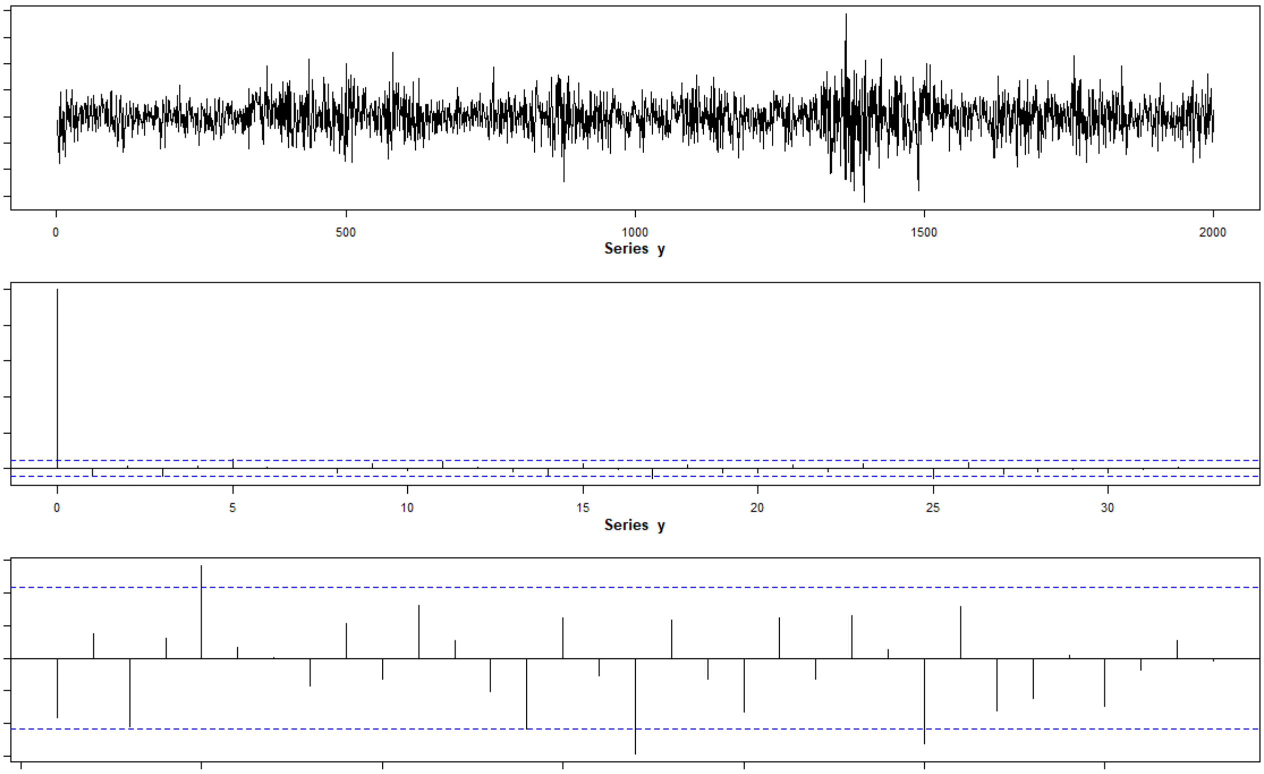


Figure 24: Plot, ACF and PACF of Series 5 Data

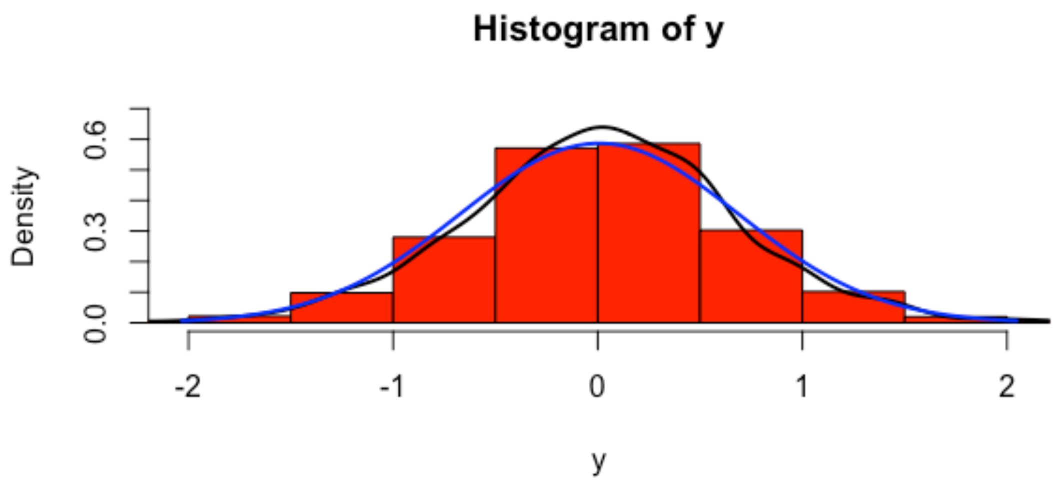


Figure 25: Histogram of Series 5 Data

```
> mean(y)
[1] 0.0071755
> sd(y)
[1] 0.6797896
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 0
> #Normality
> skewness(y)
[1] -0.08744834
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 4.36924
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.99047, p-value = 3.456e-10

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 33.151, df = 20, p-value = 0.03248
```

Figure 26: Formal Analysis of Series 5 Data with R

SERIES 6

Original

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are stationary through the mean and the variance. So, no transformation is needed.
- **Normality:** even though through the histogram we cannot say for sure if the data is normal, we can definitely say that they are not through the Shapiro test (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

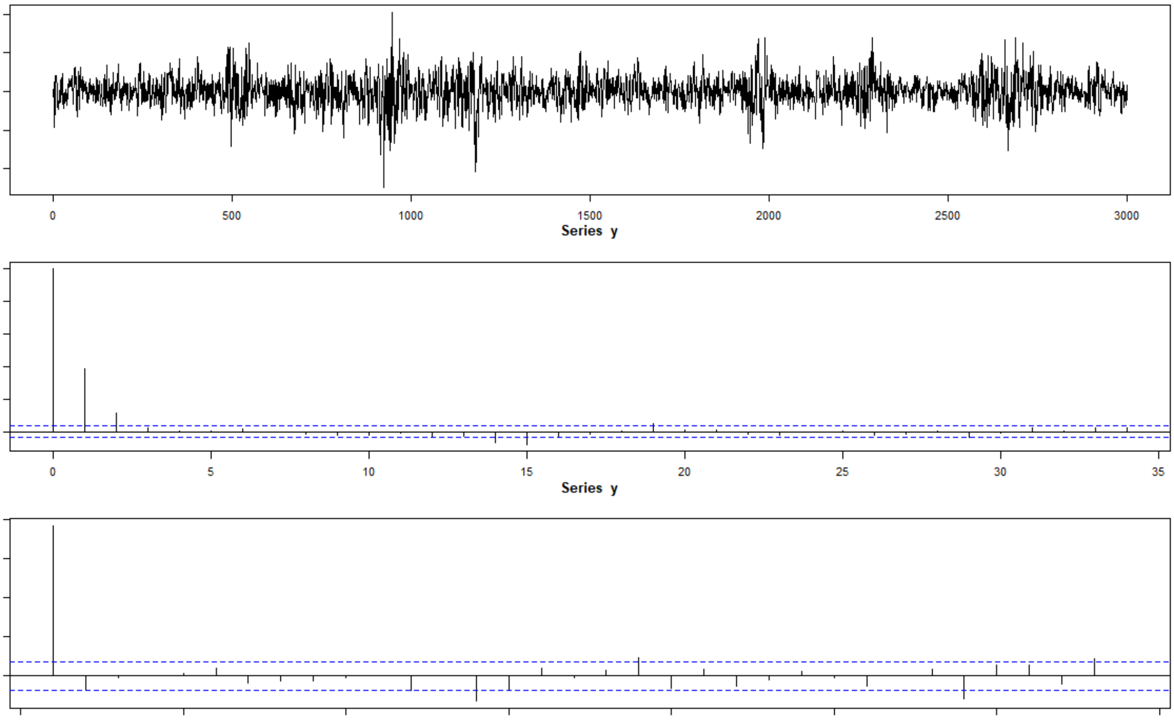


Figure 27: Plot, ACF and PACF of Series 6 Data

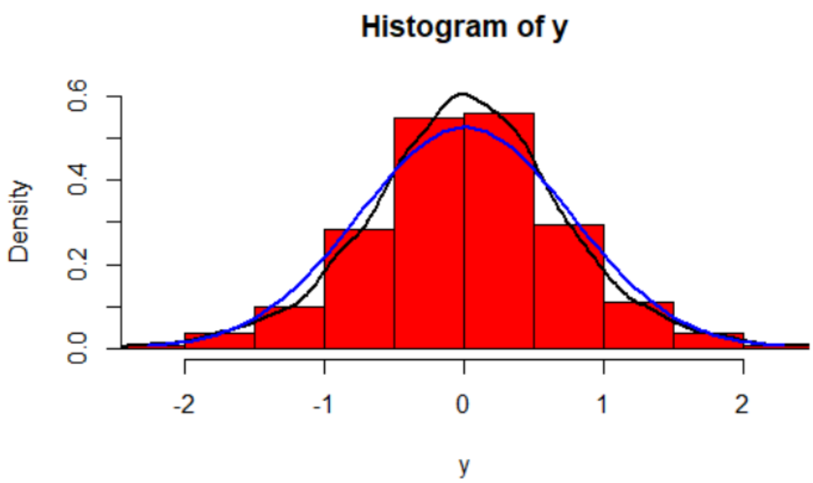
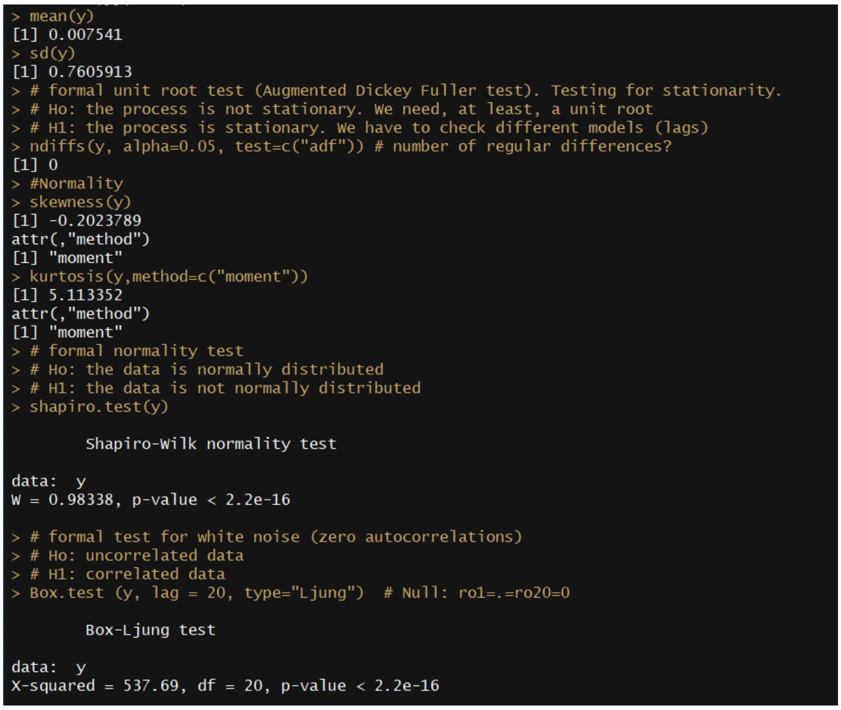


Figure 28: Histogram of Series 6 Data



```
> mean(y)
[1] 0.007541
> sd(y)
[1] 0.7605913
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 0
> #Normality
> skewness(y)
[1] -0.2023789
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 5.113352
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.98338, p-value < 2.2e-16

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 537.69, df = 20, p-value < 2.2e-16
```

Figure 29: Formal Analysis of Series 6 Data with R

SERIES 7

Original

- **Stationary:** both through the time series plot and the ADF test, we can verify that our data are not stationary through the mean. So, one transformation is needed and specifically one difference.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is not normal (p-value below 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value below 0.05). Another factor that led us to this conclusion is the distance of the mean from 0, which is one of the 3 factors that define a White Noise.
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

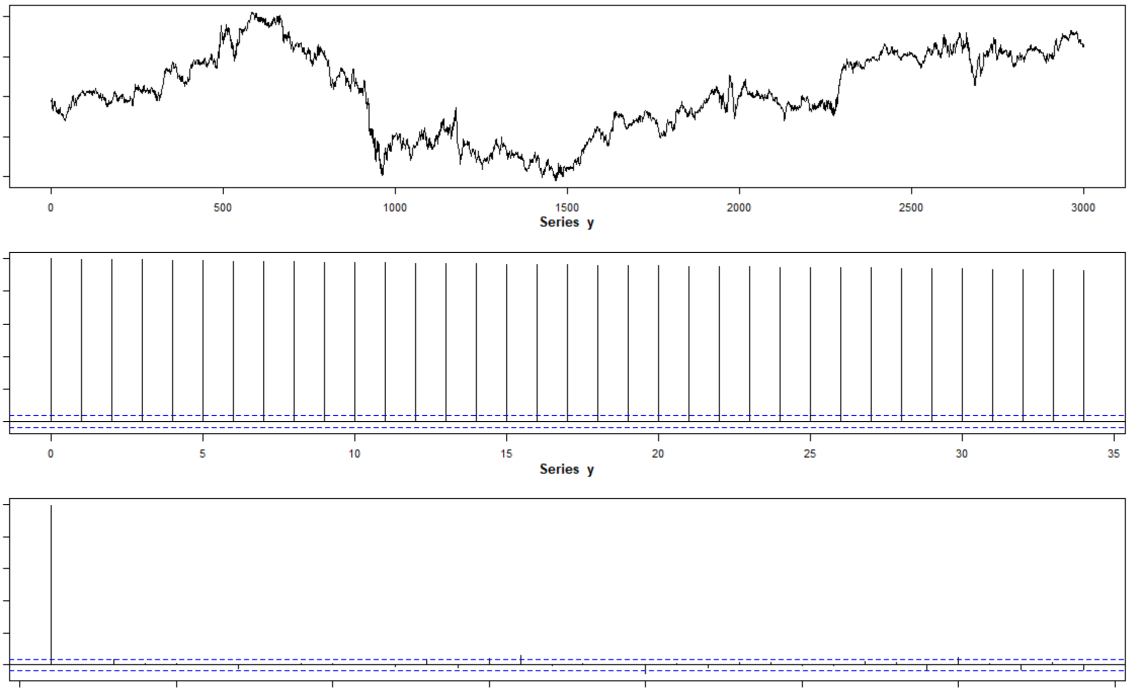


Figure 30: Plot, ACF and PACF of Series 7 Data

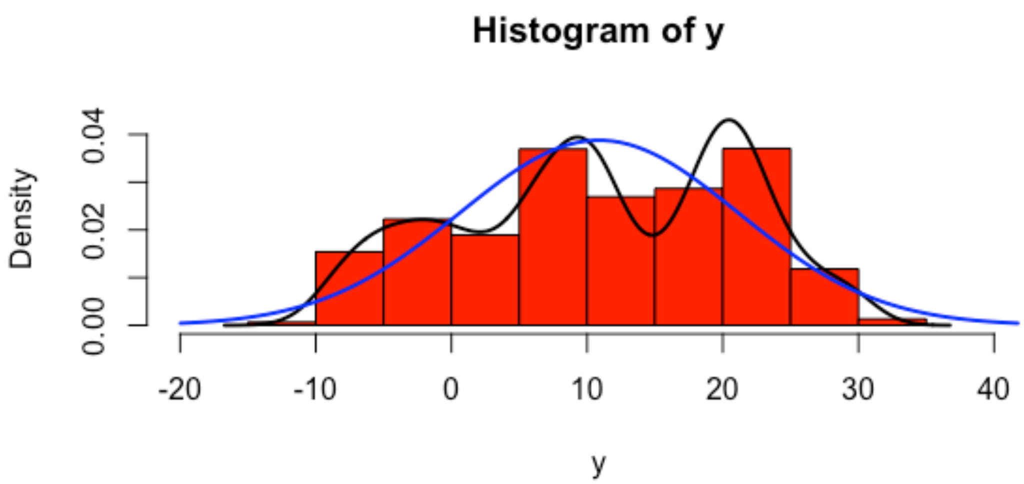


Figure 31: Histogram of Series 7 Data

```
> mean(y)
[1] 10.88196
> sd(y)
[1] 10.28588
> # formal unit root test (Augmented Dickey Fuller test). Testing for stationarity.
> # Ho: the process is not stationary. We need, at least, a unit root
> # H1: the process is stationary. We have to check different models (lags)
> ndiffs(y, alpha=0.05, test=c("adf")) # number of regular differences?
[1] 1
> #Normality
> skewness(y)
[1] -0.1702387
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 2.00975
attr(,"method")
[1] "moment"
> # formal normality test
> # Ho: the data is normally distributed
> # H1: the data is not normally distributed
> shapiro.test(y)

        Shapiro-Wilk normality test

data:  y
W = 0.96711, p-value < 2.2e-16

> # formal test for white noise (zero autocorrelations)
> # Ho: uncorrelated data
> # H1: correlated data
> Box.test (y, lag = 20, type="Ljung")  # Null: ro1=.=ro20=0

        Box-Ljung test

data:  y
X-squared = 57372, df = 20, p-value < 2.2e-16
```

Figure 32: Formal Analysis of Series 7 Data with R

Transformed

- **Stationary:** by applying one difference at our original data, we can now observe through both the plots and the ADF test, that our data is finally stationary. So, no further transformation is needed at this point.
- **Normality:** both through the histogram and the Shapiro test, we can definitely say that our data is normal (p-value above 0.05).
- **White Noise:** through the ACF and PACF plots, we believe that our data is not White Noise and we verify that by using the Box Ljung test (p-value above 0.05).
    - As the data is not WN, they cannot be SWN nor GWN.
    - As the data is not WN and by observing the spikes of the ACF and PACF, we can say that there is a linear model for our data.
    - As the data is originally not WN, we cannot explore the squared values in order to detect if there is a non-linear model, and more specifically a quadratic one.

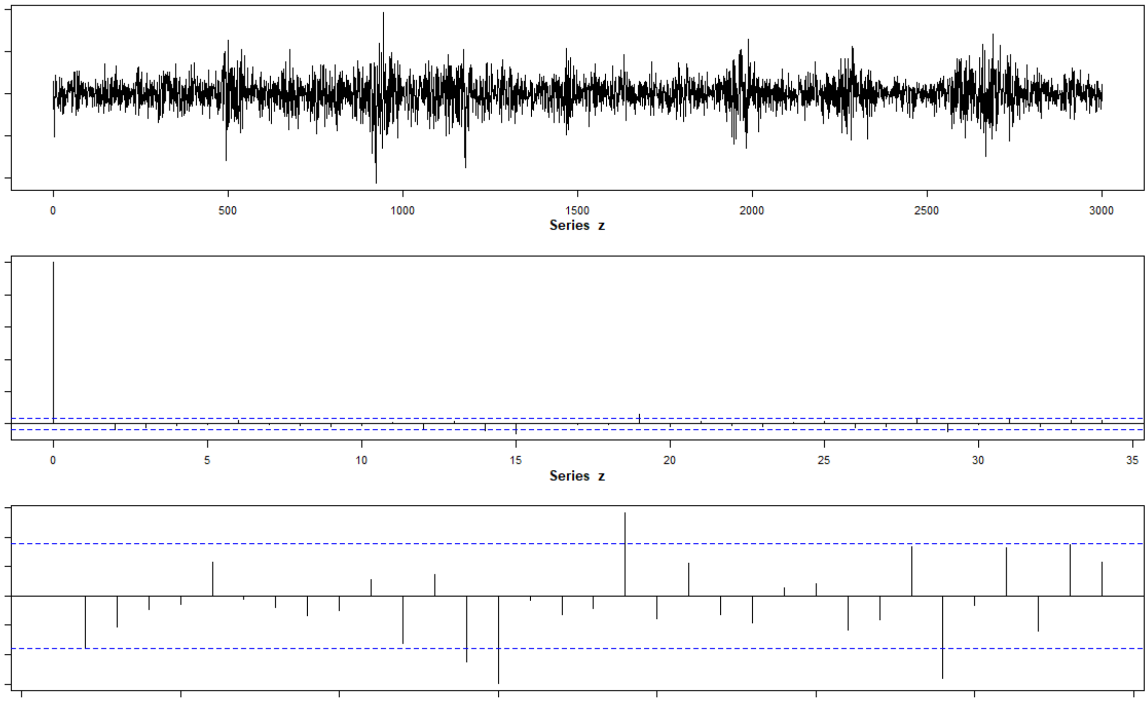


Figure 33: Plot, ACF and PACF of Series 7 Transformed Data

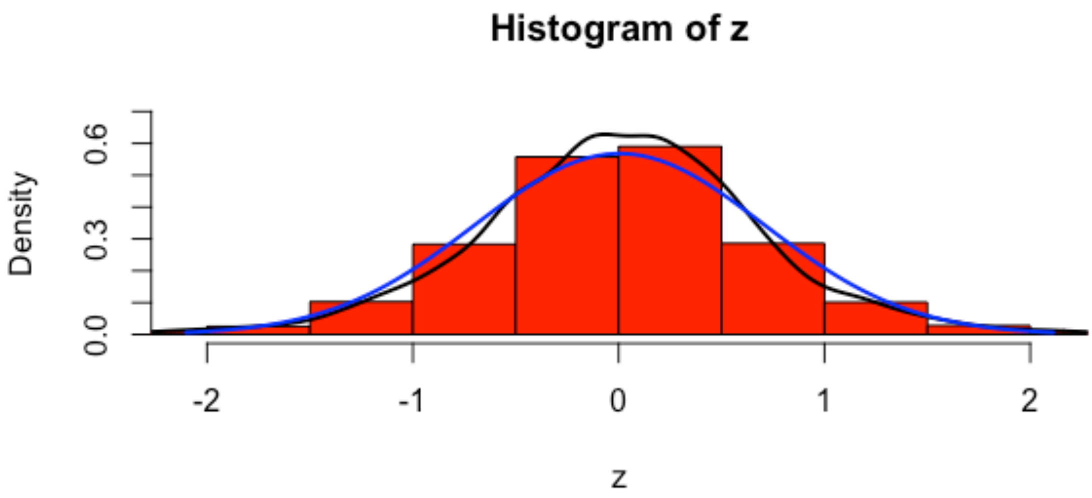


Figure 34: Histogram of Series 7 Transformed Data

```
> mean(z)
[1] 0.004649883
> sd(z)
[1] 0.7023493
> ndiffs(z, alpha=0.05, test=c("adf"))
[1] 0
> Box.test (z, lag = 20, type="Ljung")

        Box-Ljung test

data:  z
X-squared = 40.012, df = 20, p-value = 0.004978

> skewness(y)
[1] -0.1702387
attr(,"method")
[1] "moment"
> kurtosis(y,method=c("moment"))
[1] 2.00975
attr(,"method")
[1] "moment"
> shapiro.test(z)

        Shapiro-Wilk normality test

data:  z
W = 0.98584, p-value < 2.2e-16
```

Figure 35: Formal Analysis of Series 7 Transformed Data with R