HW2_Fall2020

Rong Fan

9/20/2020

Variogram function and its applications

```
# Define the variogram function
variogram_func <- function(x, lag) {
  x <- as.matrix(x) # Make sure the x is a vector. It represents the observations of y_t.
  Lag <- NULL
  var_k <- NULL
  vario <- NULL
  for (k in 1:lag) {
    Lag[k] <- k
    var_k[k] <- sd(diff(x, k))^2
    vario[k] <- var_k[k] / var_k[1]
  }
  return(as.data.frame(cbind(Lag, vario)))
}
```

Variograms of chemical process viscosity data (Text: Fig 2.17, page 45)

```
library(readxl)
viscosity <- read_excel("AppendixB_datafile.xls", skip = 2, sheet = "B.3-VISC")
```

```
## New names:
## * `Time Period` -> `Time Period...1`
## * Reading -> Reading...2
## * `Time Period` -> `Time Period...3`
## * Reading -> Reading...4
## * `Time Period` -> `Time Period...5`
## * ...
```

```
readings <- na.omit(c(viscosity$Reading...2, viscosity$Reading...4,
                      viscosity$Reading...6, viscosity$Reading...8))
head(readings)
```

```
## [1] 86.7418 85.3195 84.7355 85.1113 85.1487 84.4775
```

```
class(readings)
```

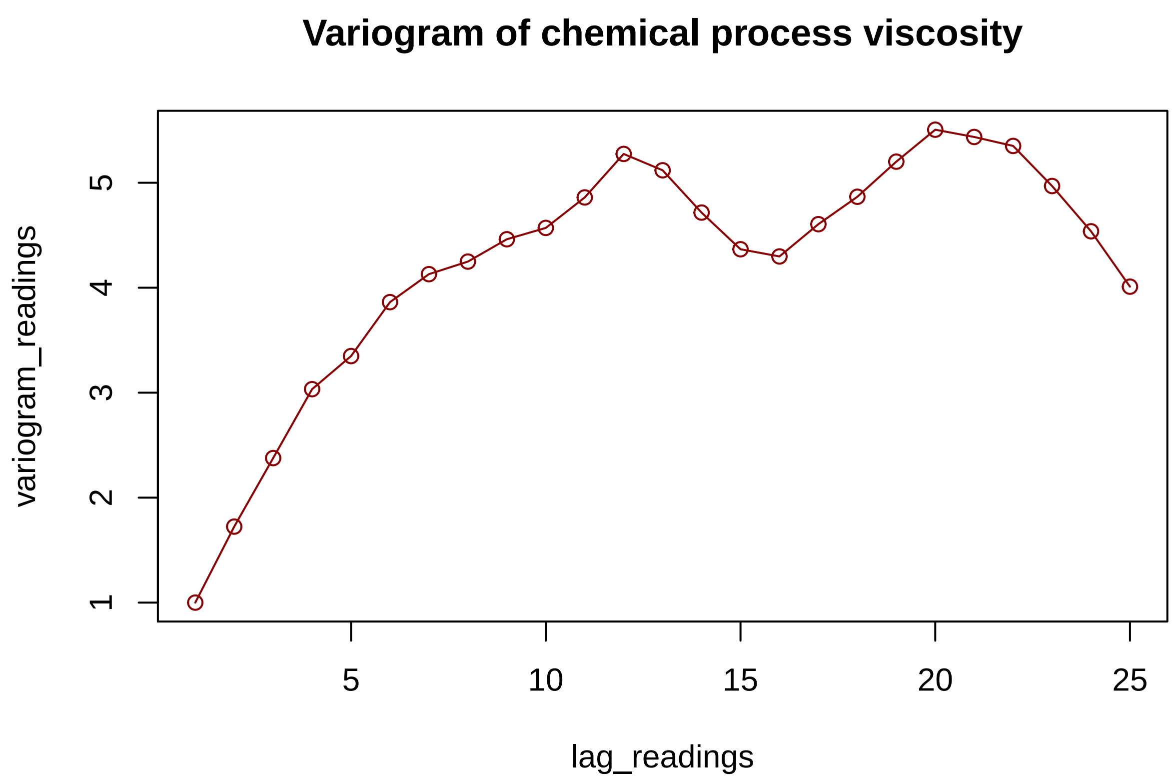
```
## [1] "numeric"
```

```
x <- readings
lag_length <- 25
lag_readings <- 1:lag_length
```

```r
z <- variogram_func(x, lag_length)
variogram_readings <- z$vario
variogram_readings
```

```
##  [1] 1.000000 1.723942 2.376856 3.033604 3.348690 3.862875 4.129339 4.248845
##  [9] 4.461848 4.569809 4.860165 5.274703 5.119000 4.715869 4.366605 4.298101
## [17] 4.605405 4.866596 5.201105 5.505501 5.436292 5.350848 4.968960 4.537469
## [25] 4.010016
```

```r
plot(lag_readings, variogram_readings, type = "o", col = "dark red",
     main = "Variogram of chemical process viscosity")
```



Variograms of Whole Foods Market stock Price (Text: Fig 2.19, page 46)

```r
library(readxl)
stock <- read_excel("AppendixB_datafile.xls", skip = 3, sheet = "B.7-WFMI")
```

```
## New names:
## * Date -> Date...1
## * Dollars -> Dollars...2
## * Date -> Date...3
## * Dollars -> Dollars...4
## * Date -> Date...5
## * ...
```

```r
price <- na.omit(c(stock$Dollars...2, stock$Dollars...4,
                   stock$Dollars...6, stock$Dollars...8, stock$Dollars...10))
head(price)
```

```
## [1] 28.05 28.23 26.25 25.41 26.25 26.03
```

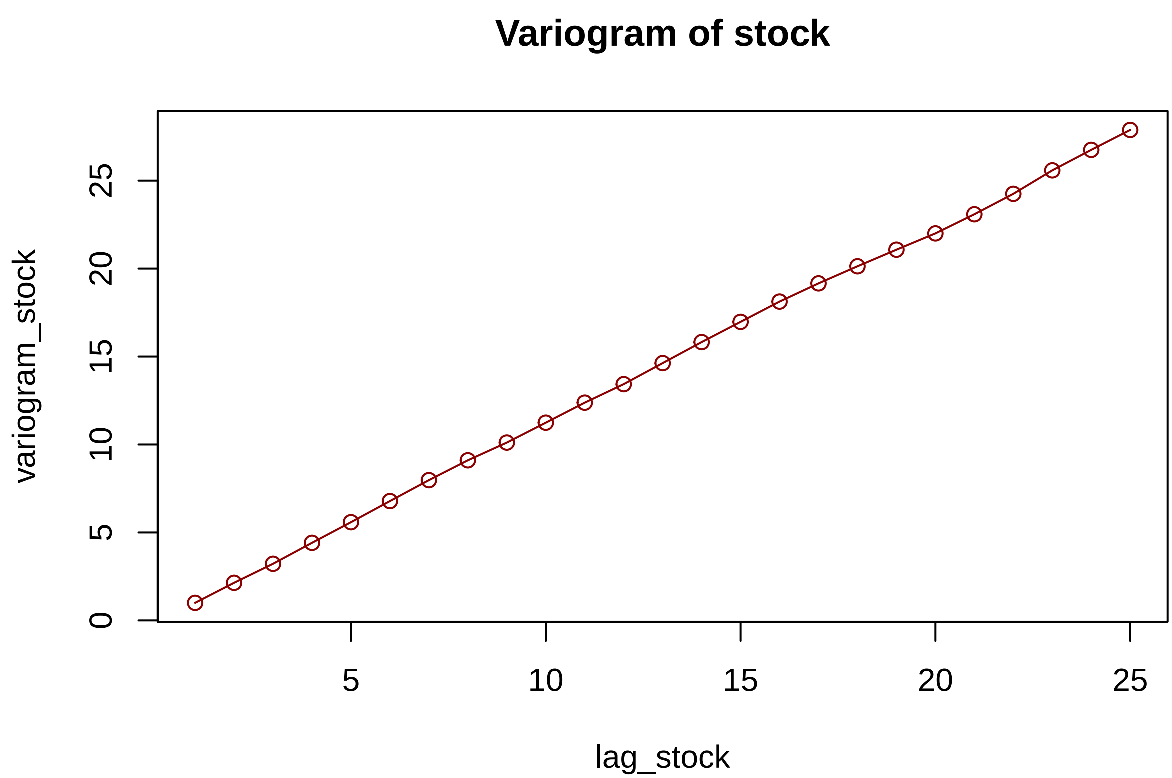
```r
class(price)
```

```
## [1] "numeric"
```

```r
x <- price
lag_length <- 25
lag_stock <- 1:lag_length
z <- variogram_func(x, lag_length)
variogram_stock <- z$vario
variogram_stock
```

```
##  [1]  1.000000  2.139660  3.219850  4.411881  5.583359  6.784479  7.974550
##  [8]  9.102236 10.109781 11.240052 12.378525 13.431489 14.627292 15.821995
## [15] 16.972685 18.123384 19.159940 20.130744 21.074192 21.999525 23.088225
## [22] 24.249945 25.582925 26.749122 27.881126
```

```r
plot(lag_stock, variogram_stock, type = "o", col = "dark red",
     main = "Variogram of stock")
```

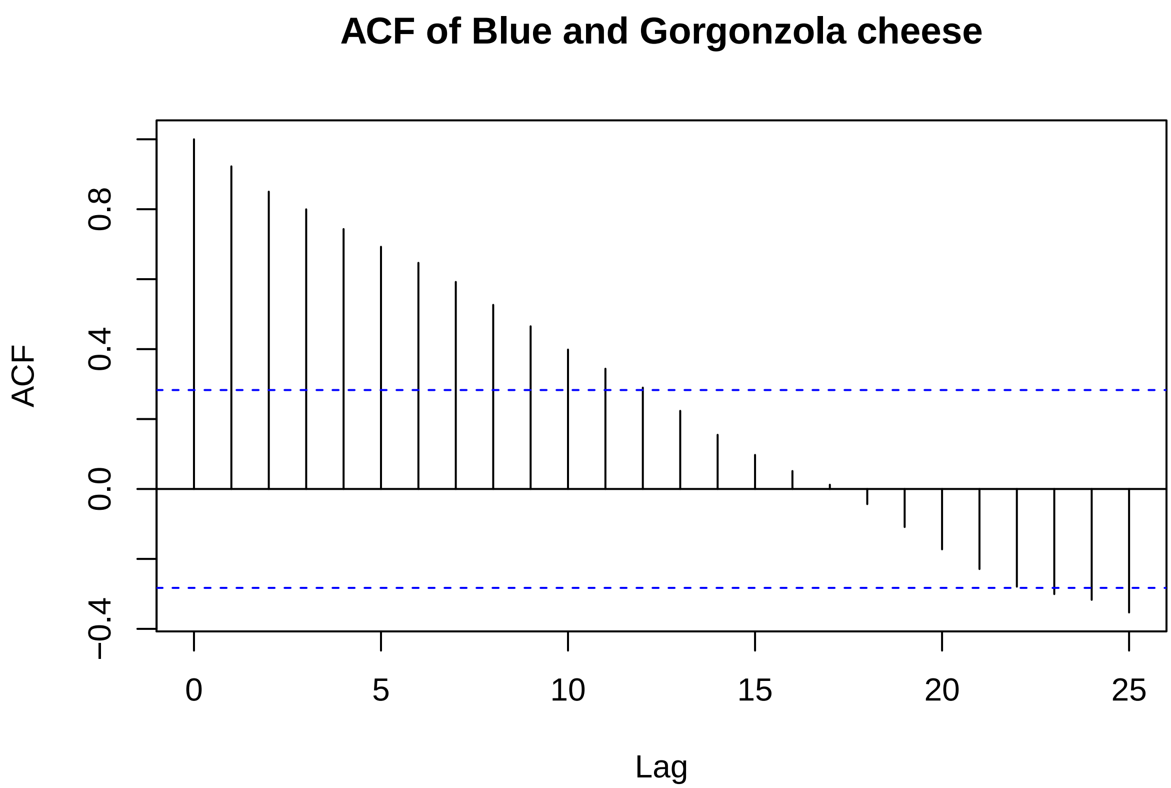


HW2

R version

```
library(readxl)
Blue <- read_excel("AppendixB_datafile.xls",
  sheet = "B.4-BLUE"
)
```

```
x <- acf(Blue[, 2], lag.max = 25, type = "correlation",
         main = "ACF of Blue and Gorgonzola cheese") # ACF plot
```



2.2 a)

```
x$acf[1:26] # values of ACFs for the first 25 lags
```

```
##  [1]  1.00000000  0.92255797  0.85024468  0.79968986  0.74328987  0.69263229
##  [7]  0.64664768  0.59209915  0.52647671  0.46521790  0.39867153  0.34387841
## [13]  0.28989096  0.22336880  0.15486809  0.09742316  0.05132157  0.01216789
## [19] -0.04327517 -0.10876777 -0.17264515 -0.22894671 -0.27937815 -0.30055353
## [25] -0.31694514 -0.35312179
```

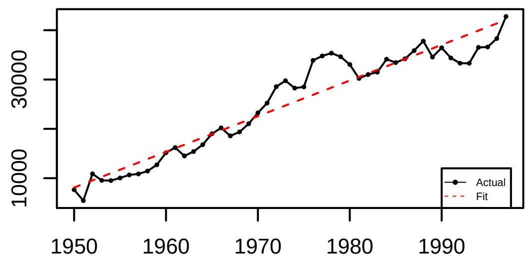
```
x <- Blue$Production
lag_length <- length(x) / 4
lag_cheese <- 1:lag_length
z <- variogram_func(x, lag_length)
variogram_cheese <- z$vario
variogram_cheese
```

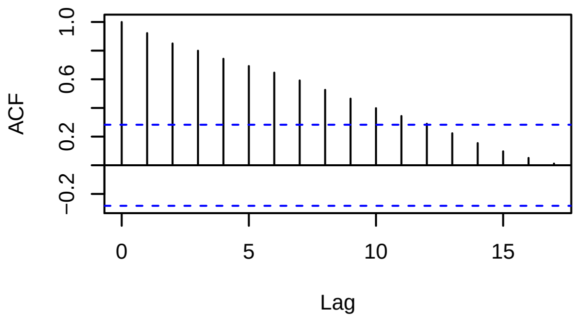
```
plot(lag_cheese, variogram_cheese, type = "o", col = "dark red",
     main = "Variogram of Blue and Gorgonzola cheese")
```

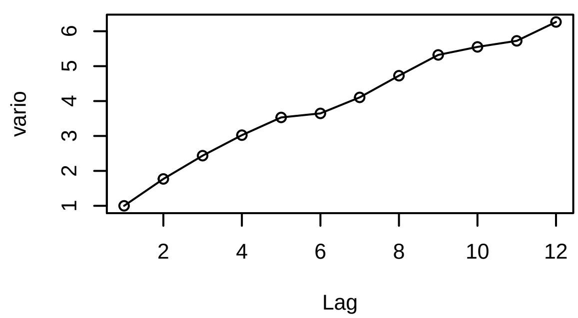
More general, let's get all graphs related to the Bule and Gorgonzola cheese data. You can ignore the code below if you don't want to get all the outputs in one function.

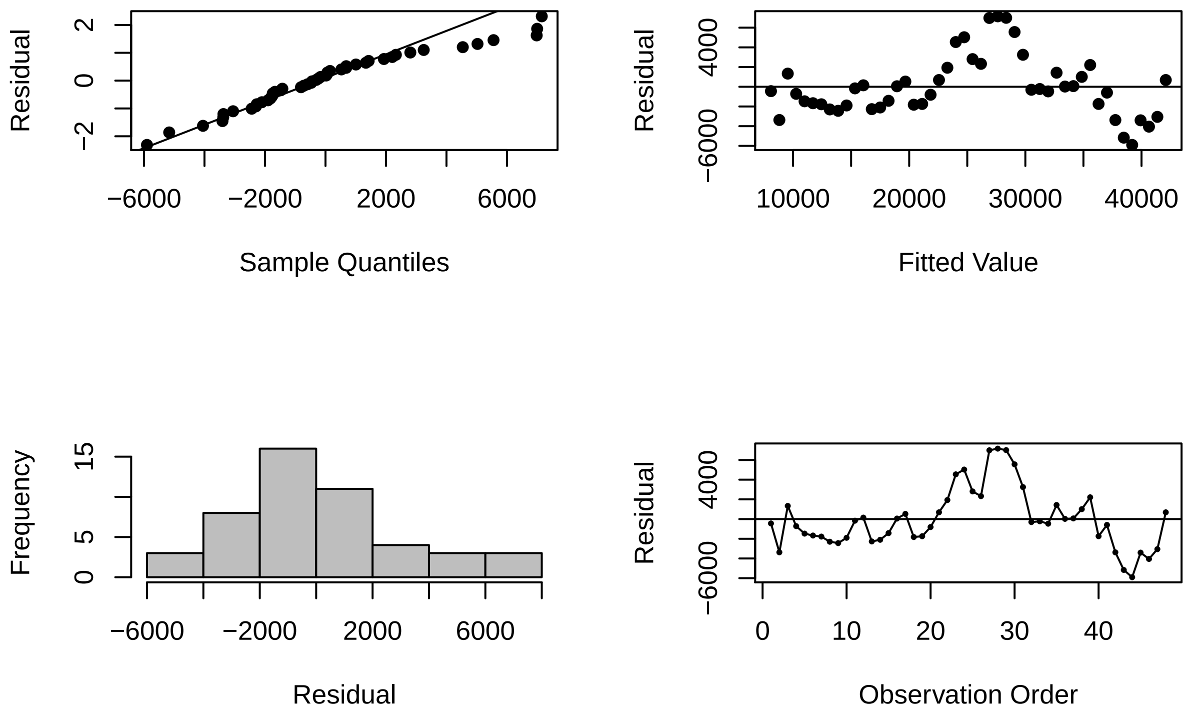
```
time.series.graphics <- function(data, title = "", xlab = "", ylab = "") {
  # Use the variogram_func() defined above
  # or include the definition of the variogram function within the chunk so that the function
  # time.series.graphics() can be used independently.
  # Define the variogram function
  variogram <- function(x, lag) {
    x <- as.matrix(x)
    Lag <- NULL
    var_k <- NULL
    vario <- NULL
    for (k in 1:lag) {
      Lag[k] <- k
      var_k[k] <- sd(diff(x, k))^2
      vario[k] <- var_k[k] / var_k[1]
    }
    return(as.data.frame(cbind(Lag, vario)))
  }
  # End of variogram function

  layout(mat = matrix(c(0, 1, 1, 0, 2, 2, 3, 3), nrow = 2, byrow = T))

  fit.data <- lm(data[, 2] ~ data[, 1])
  plot(data, type = "l", ylab = ylab, xlab = xlab, main = title)
  points(data, pch = 16, cex = .5)
  lines(data[, 1], fit.data$fitted.values, col = "red", lty = 2)
  legend(1990, 12000, c("Actual", "Fit"), pch = (c(16, NA)), lwd = c(.5, .5),
         lty = c(1, 2), cex = .55, col = c("black", "red"))

  acf.data <- acf(data[, 2], lag.max = 17, type = "correlation", main = "")

  data.vario <- variogram(data[, 2], length(data[, 2]) / 4)

  plot(data.vario, type = "l")
  points(data.vario)

  par(mfrow = c(2, 2), oma = c(0, 0, 0, 0))
  qqnorm(fit.data$residuals, datax = T, pch = 16, xlab = "Residual", main = "")
  qqline(fit.data$residuals, data = T)
  plot(fit.data$fitted.values, fit.data$residuals, pch = 16,
       xlab = "Fitted Value", ylab = "Residual")
  abline(h = 0)
  hist(fit.data$residuals, col = "grey", xlab = "Residual", main = "")
  plot(fit.data$residuals, type = "l", xlab = "Observation Order", ylab = "Residual")
  points(fit.data$residuals, pch = 16, cex = .5)
  abline(h = 0)
}
```

```r
newBlue <- as.data.frame(Blue)
time.series.graphics(newBlue)
```









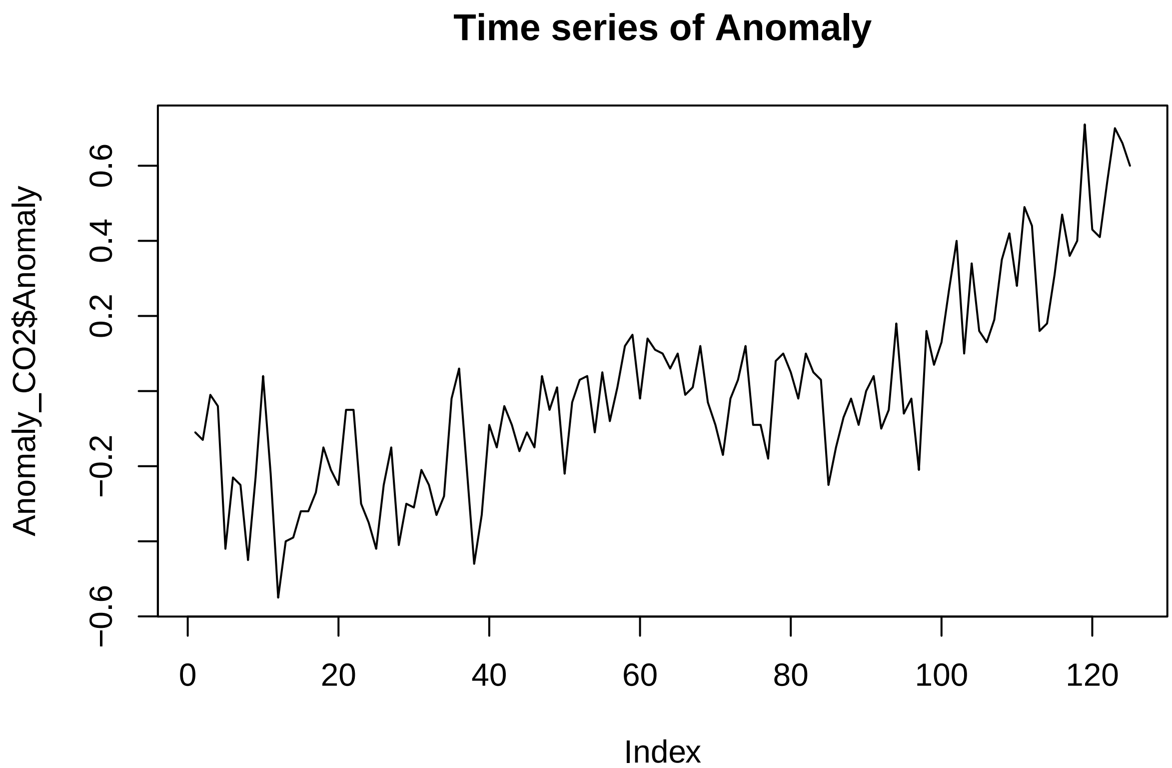
The time series plot, ACF and variogram show that the US production of blue and gorgonzola cheeses data is non-stationary.

```r
library(readxl)
Anomaly_CO2 <- read_excel("AppendixB_datafile.xls", skip = 0, sheet = "B.6-GSAT-CO2")
head(Anomaly_CO2)
```

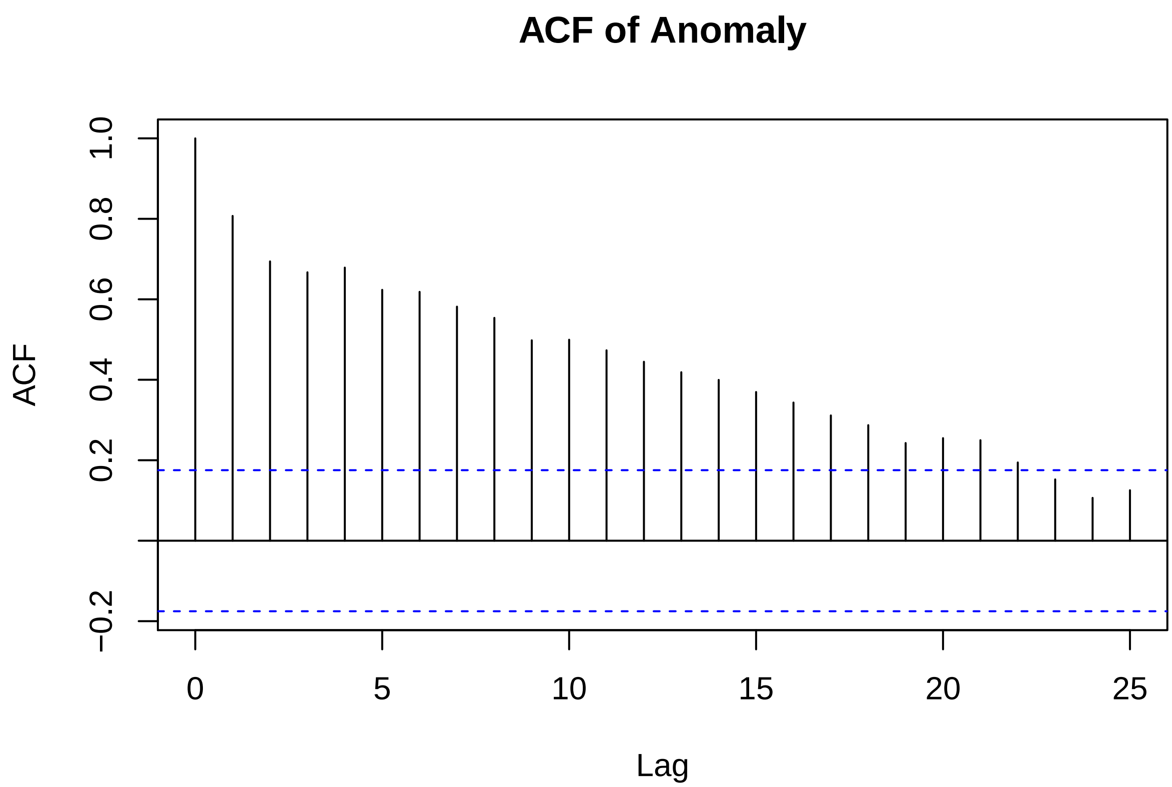
2.4.1 ACF and variogram of Anomaly

```
## # A tibble: 6 x 3
##    Year Anomaly   CO2
##   <dbl>   <dbl> <dbl>
## 1  1880   -0.11  291.
## 2  1881   -0.13  291.
## 3  1882   -0.01  292.
## 4  1883   -0.04  292.
## 5  1884   -0.42  293.
## 6  1885   -0.23  293
```

```r
plot(Anomaly_CO2$Anomaly, type = "l", main = "Time series of Anomaly")
```




```
x <- acf(Anomaly_CO2$Anomaly, lag.max = 25, type = "correlation",
         main = "ACF of Anomaly") # ACF plot
```



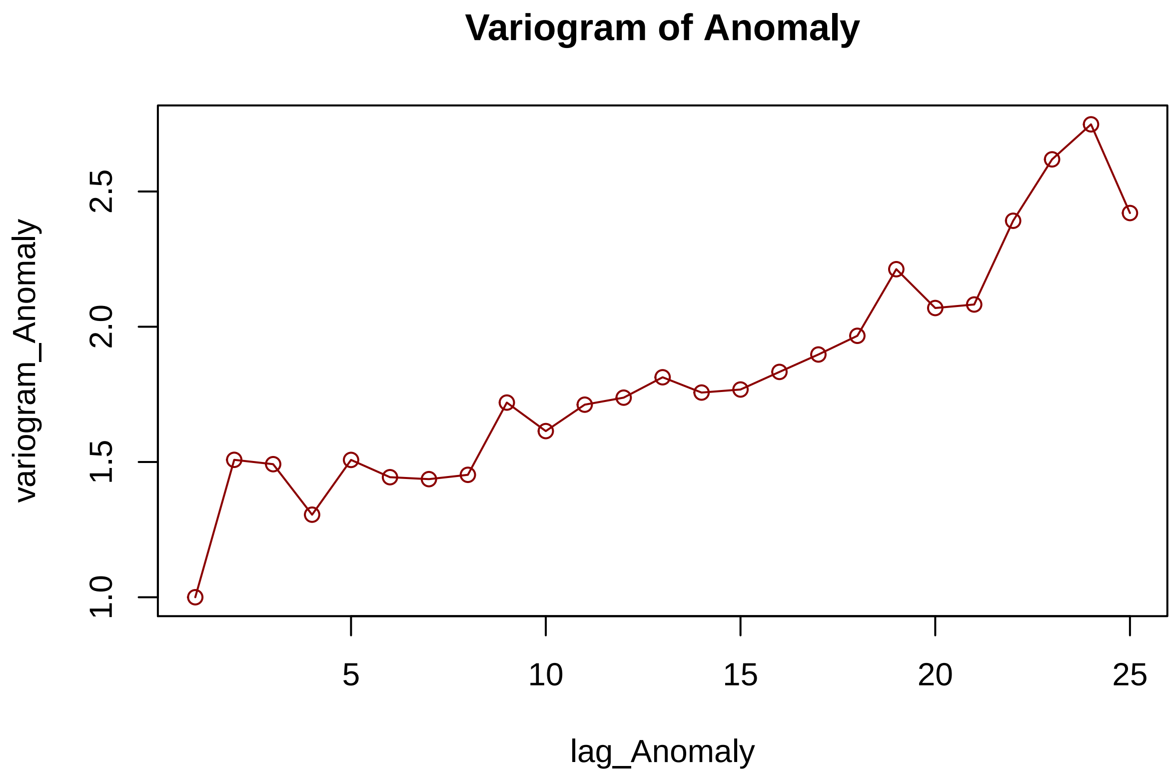

```
x$acf[1:26] # values of ACFs for the first 25 lags
```

```
##  [1] 1.0000000 0.8071948 0.6942248 0.6670964 0.6787528 0.6235193 0.6184973
##  [8] 0.5817489 0.5538216 0.4980878 0.4995987 0.4732532 0.4447530 0.4187977
## [15] 0.3996724 0.3696274 0.3434193 0.3114057 0.2871918 0.2427561 0.2549324
## [22] 0.2498699 0.1944899 0.1524225 0.1064507 0.1255333
```

```
x <- Anomaly_CO2$Anomaly
lag_length <- 25
lag_Anomaly <- 1:lag_length
z <- variogram_func(x, lag_length)
variogram_Anomaly <- z$vario
variogram_Anomaly
```

```
##  [1] 1.000000 1.508014 1.491944 1.305226 1.507647 1.443786 1.436555 1.452578
##  [9] 1.719642 1.614282 1.711940 1.737889 1.813212 1.756689 1.767938 1.833040
## [17] 1.897329 1.966501 2.212713 2.069221 2.082261 2.391755 2.618794 2.748224
## [25] 2.420406
```

```
plot(lag_Anomaly, variogram_Anomaly, type = "o", col = "dark red",
     main = "Variogram of Anomaly")
```



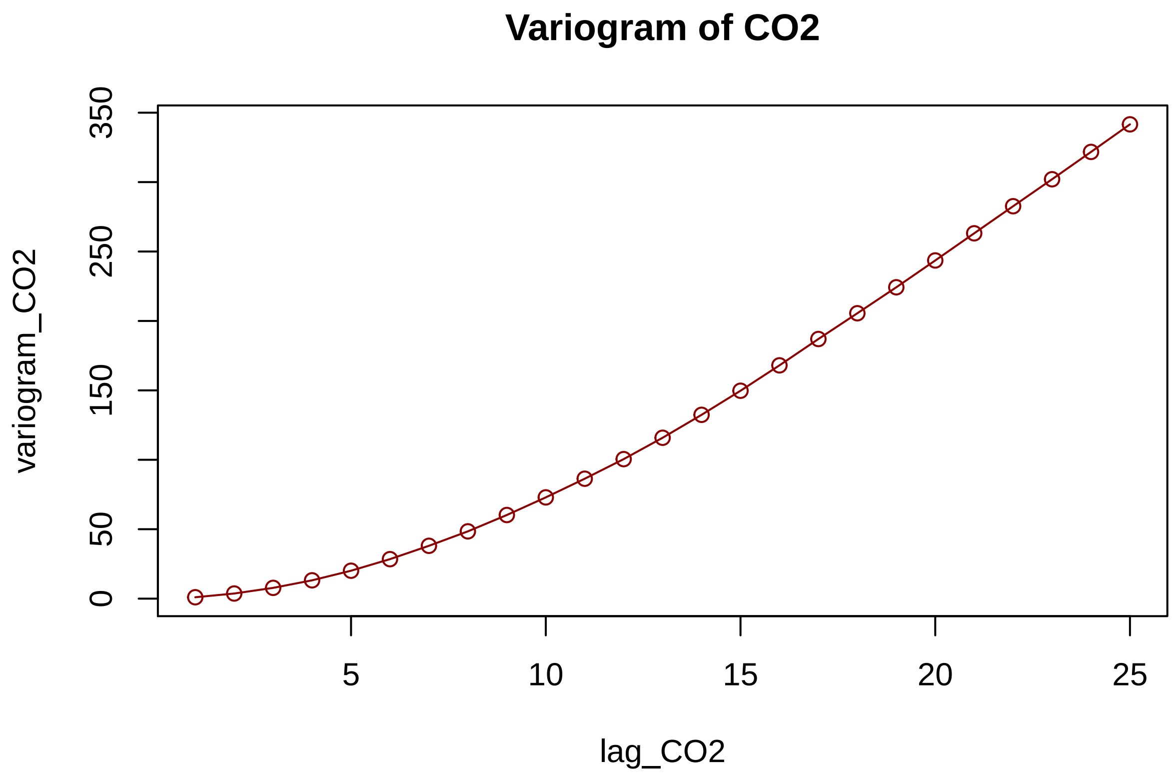
The time series plot, ACF and variogram show that the anomaly data is non-stationary.

```
x <- Anomaly_CO2$CO2
lag_length <- 25
lag_CO2 <- 1:lag_length
z <- variogram_func(x, lag_length)
variogram_CO2 <- z$vario
variogram_CO2
```

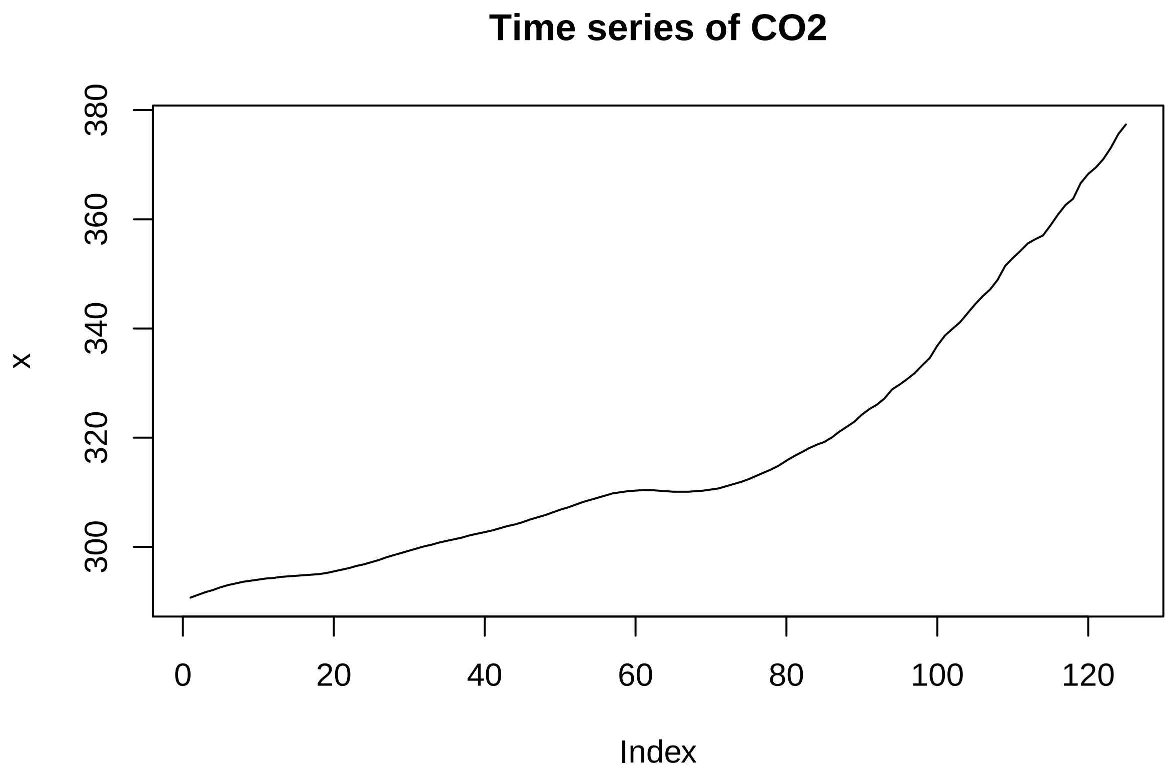
2.4.2 ACF and variogram of CO2

```
##  [1]   1.000000   3.684466   7.751564  13.189532  20.099805  28.440840
##  [7]  38.038272  48.450856  60.223840  72.931205  86.374042 100.487887
## [13] 115.864015 132.391615 149.733282 168.023418 186.988846 205.541761
## [19] 224.244712 243.576099 263.123240 282.636155 302.050303 321.739180
## [25] 341.597948
```

```
plot(lag_CO2, variogram_CO2, type = "o", col = "dark red",
     main = "Variogram of CO2")
```




```r
plot(x, type = "l", main = "Time series of CO2")
```



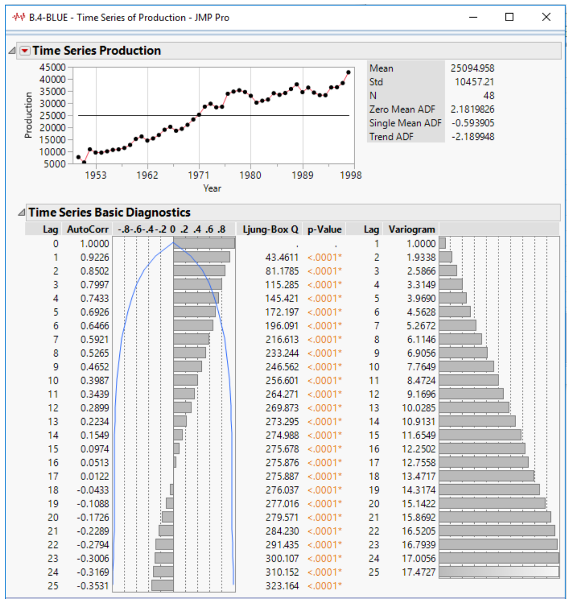
The time series plot, ACF and variogram show that the CO2 data is non-stationary.

JMP version

2.2 a) ACF and variogram of Blue and Gorgonzola cheese

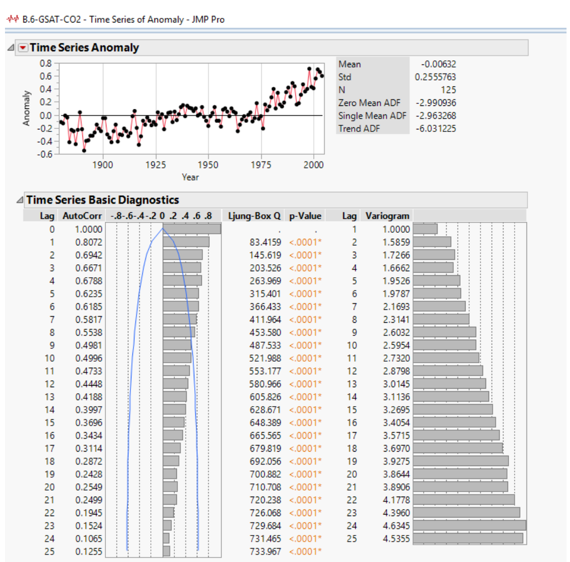
2.4.1 ACF and variogram of Anomaly

2.4.2 ACF and variogram of CO2



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Variogram
0	1.0000	.	.	1	1.0000
1	0.9226	43.4611	<.0001*	2	1.9338
2	0.8502	81.1785	<.0001*	3	2.5866
3	0.7997	115.285	<.0001*	4	3.3149
4	0.7433	145.421	<.0001*	5	3.9690
5	0.6926	172.197	<.0001*	6	4.5628
6	0.6466	196.091	<.0001*	7	5.2672
7	0.5921	216.613	<.0001*	8	6.1146
8	0.5265	233.244	<.0001*	9	6.9056
9	0.4652	246.562	<.0001*	10	7.7649
10	0.3987	256.601	<.0001*	11	8.4724
11	0.3439	264.271	<.0001*	12	9.1696
12	0.2899	269.873	<.0001*	13	10.0285
13	0.2234	273.295	<.0001*	14	10.9131
14	0.1549	274.988	<.0001*	15	11.6549
15	0.0974	275.678	<.0001*	16	12.2502
16	0.0513	275.876	<.0001*	17	12.7558
17	0.0122	275.887	<.0001*	18	13.4717
18	-0.0433	276.037	<.0001*	19	14.3174
19	-0.1088	277.016	<.0001*	20	15.1422
20	-0.1726	279.571	<.0001*	21	15.8692
21	-0.2289	284.230	<.0001*	22	16.5205
22	-0.2794	291.435	<.0001*	23	16.7939
23	-0.3006	300.107	<.0001*	24	17.0056
24	-0.3169	310.152	<.0001*	25	17.4727
25	-0.3531	323.164	<.0001*		

Figure 1: ACF and Variogram of blue and gorgonzola cheese



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Variogram
0	1.0000	.	.	1	1.0000
1	0.8072	83.4159	<.0001*	2	1.5859
2	0.6942	145.619	<.0001*	3	1.7266
3	0.6671	203.526	<.0001*	4	1.6662
4	0.6788	263.969	<.0001*	5	1.9526
5	0.6235	315.401	<.0001*	6	1.9787
6	0.6185	366.433	<.0001*	7	2.1693
7	0.5817	411.964	<.0001*	8	2.3141
8	0.5538	453.580	<.0001*	9	2.6032
9	0.4981	487.533	<.0001*	10	2.5954
10	0.4996	521.988	<.0001*	11	2.7320
11	0.4733	553.177	<.0001*	12	2.8798
12	0.4448	580.966	<.0001*	13	3.0145
13	0.4188	605.826	<.0001*	14	3.1136
14	0.3997	628.671	<.0001*	15	3.2695
15	0.3696	648.389	<.0001*	16	3.4054
16	0.3434	665.565	<.0001*	17	3.5715
17	0.3114	679.819	<.0001*	18	3.6970
18	0.2872	692.056	<.0001*	19	3.9275
19	0.2428	700.882	<.0001*	20	3.8644
20	0.2549	710.708	<.0001*	21	3.8906
21	0.2499	720.238	<.0001*	22	4.1778
22	0.1945	726.068	<.0001*	23	4.3960
23	0.1524	729.684	<.0001*	24	4.6345
24	0.1065	731.465	<.0001*	25	4.5355
25	0.1255	733.967	<.0001*		

Figure 2: ACF and Variogram of Anomaly

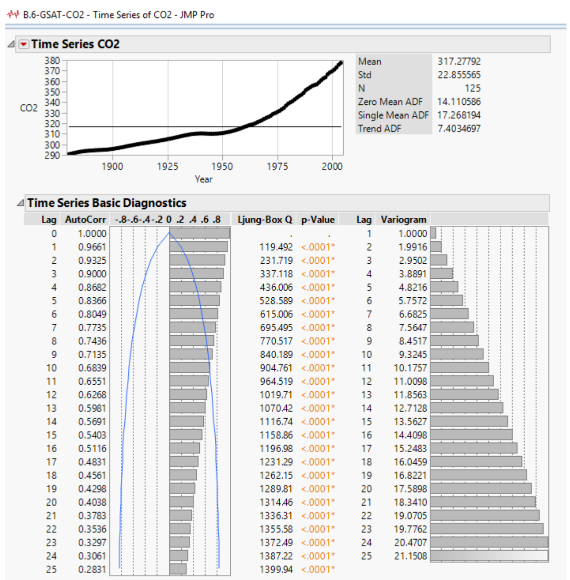


Figure 3: ACF and Variogram of CO2