STATISTICS 3N03/3J04

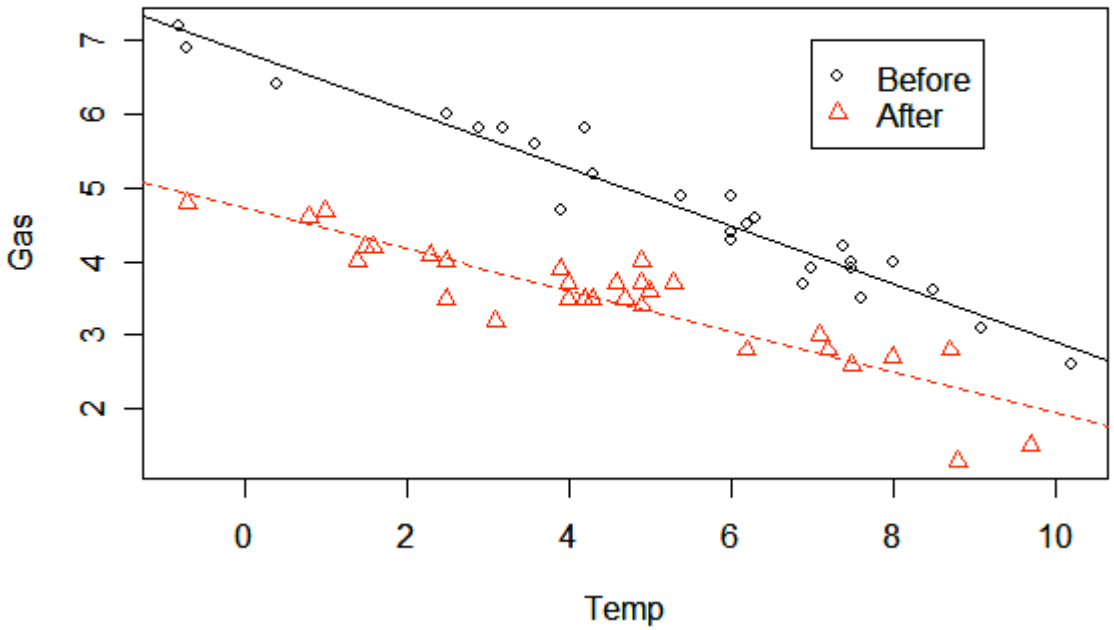
Test 1b – 2006-10-05 – Solutions

Question 1. **[12 marks for appropriate conclusions supported by graphs.]**

The gas consumption required to heat Mr Whiteside's house appears to be linearly related to temperature, with negative slope. After the insulation was installed, the slope of this relationship was reduced. The insulation saved about 2000 ft^3 of gas per week when the mean weekly temperature was 0^o C, compared to about 1000 ft^3 per week at 10^o C.

If you inspect the whiteside data frame, the rows have evidently been sorted by Temp within Insul. This means that the time order of the observations has been lost so we can't analyze them as a time series.

```
> library(MASS)
> names(whiteside)
[1] "Insul" "Temp"  "Gas"
> levels(whiteside$Insul)
[1] "Before" "After"
> plot(Gas~Temp, whiteside, pch=as.numeric(Insul), col=as.numeric(Insul))
> abline(lm(Gas~Temp, whiteside[whiteside$Insul=="Before",]), col=1, lty=1)
> abline(lm(Gas~Temp, whiteside[whiteside$Insul=="After",]), col=2, lty=2)
> legend(7, 7, levels(whiteside$Insul), col=1:2, pch=1:2)
```



Question 2. **[12 marks for appropriate conclusions supported by graphs.]**

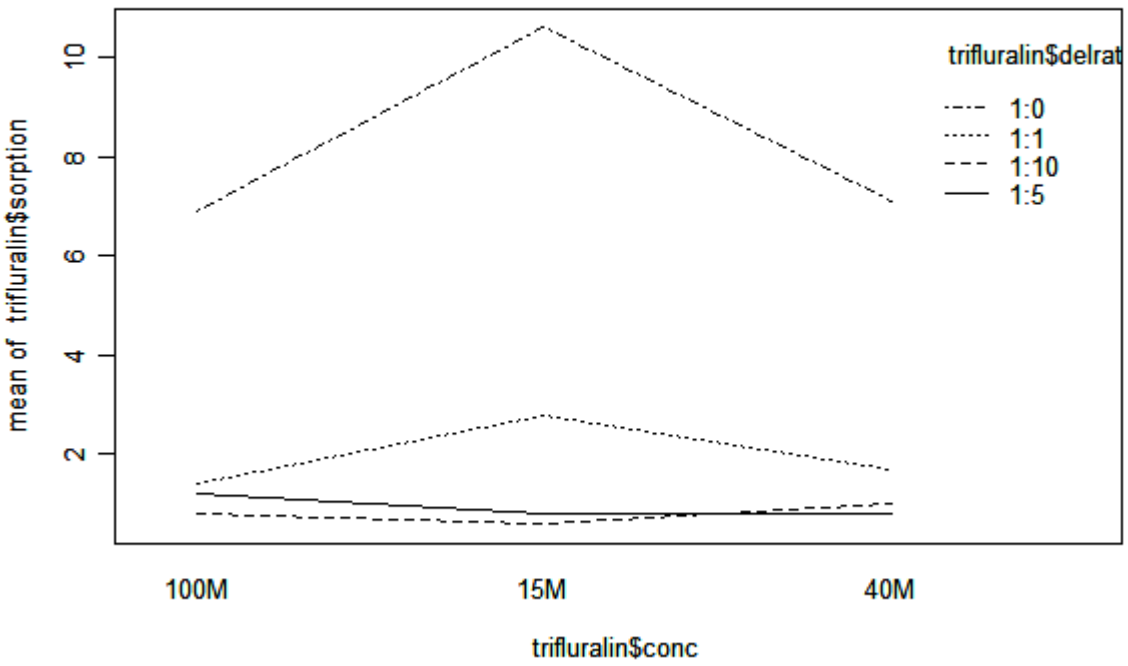
From either interaction plot, we can see that a 1:0 delivery ratio gives maximum sorption at any concentration, while sorption at a 1:1 delivery ratio is only slightly higher than at 1:5 or 1:10. When the delivery ratio is 1:0, 40M gives a much higher sorption that the higher or lower concentrations; this effect is much less when the delivery ratio is 1:1 and does not happen at 1:5 or 1:10 so we conclude that the two factors interact (the lines are not parallel). The highest sorption is achieved at delivery ratio 1:0 and concentration 15M.

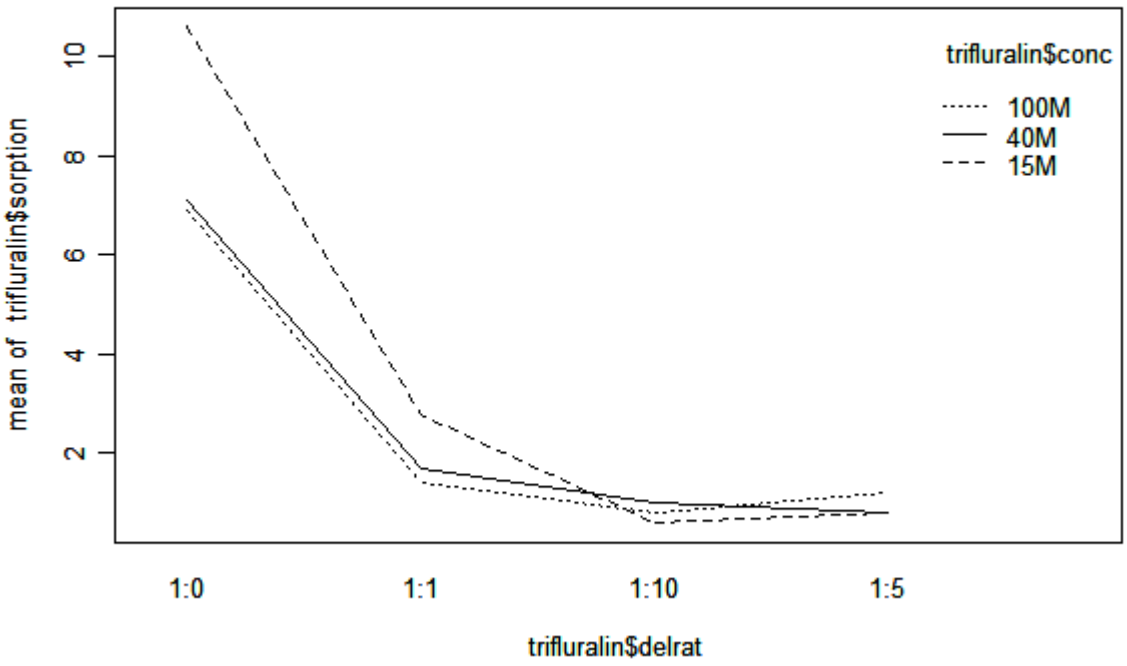
The box plots lead to the same conclusion, but also show the variation. The variability of sorption rises with the mean; plotting sorption on a log scale makes the variation more consistent over the different experimental conditions.

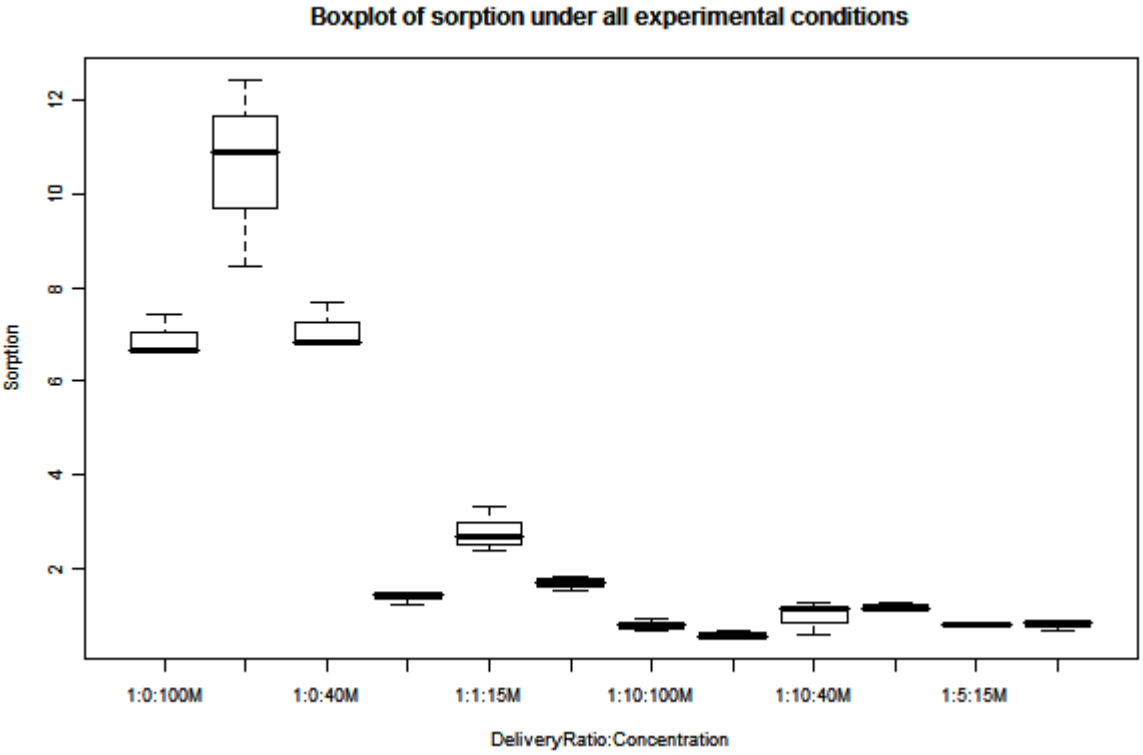
Box plots of each factor separately would also be interesting, but because both factors have strong effects they will not show the differences so clearly.

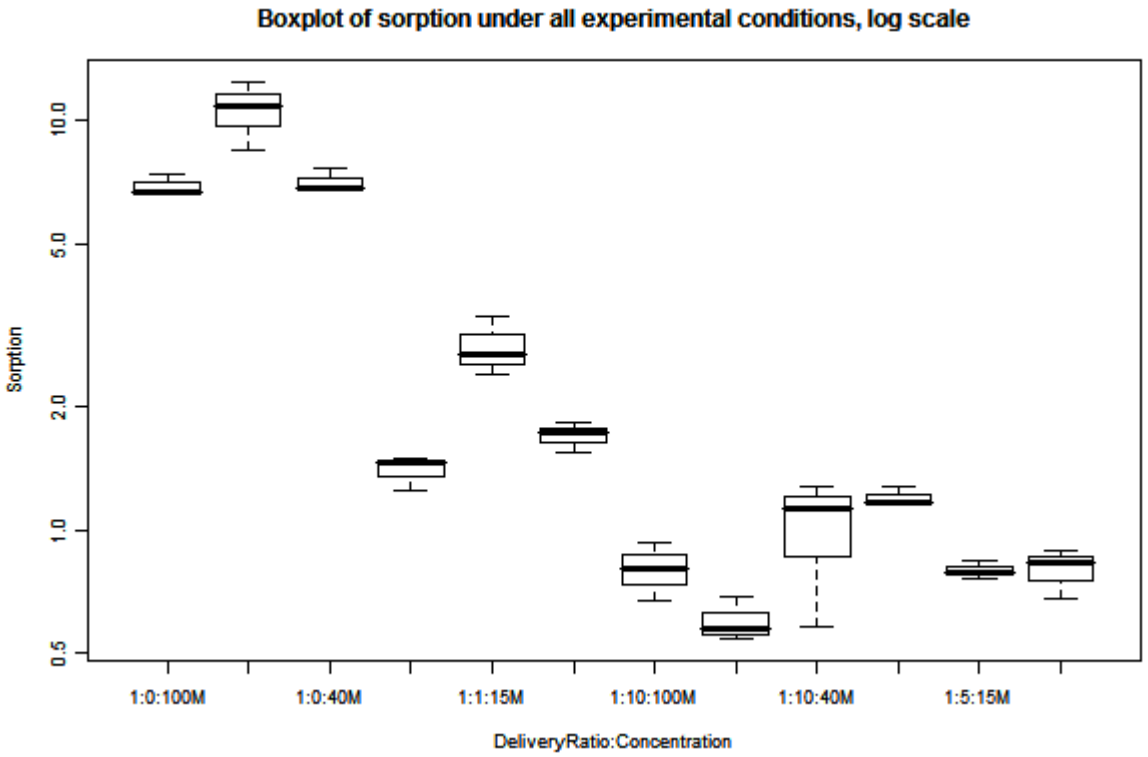
```
> trifluralin <- data.frame(sorption=c(10.90, 8.47, 12.43, 3.33, 2.40, 2.67, 0.79, 0.76,
0.84, 0.54, 0.69, 0.57, 6.84, 7.68, 6.79, 1.72, 1.55, 1.82, 0.68, 0.83, 0.89, 0.58, 1.13,
1.28, 6.61, 6.66, 7.43, 1.25, 1.46, 1.49, 1.17, 1.27, 1.16, 0.93, 0.67, 0.80),
conc=rep(c("15M","40M","100M"),rep(12,3)),
delrat=rep(rep(c("1:0","1:1","1:5","1:10"), rep(3,4)),3))
> trifluralin
   sorption conc  delrat
1     10.90  15M     1:0
2      8.47  15M     1:0
3     12.43  15M     1:0
4      3.33  15M     1:1
5      2.40  15M     1:1
6      2.67  15M     1:1
7      0.79  15M     1:5
8      0.76  15M     1:5
9      0.84  15M     1:5
10     0.54  15M    1:10
11     0.69  15M    1:10
12     0.57  15M    1:10
13     6.84  40M     1:0
14     7.68  40M     1:0
15     6.79  40M     1:0
16     1.72  40M     1:1
17     1.55  40M     1:1
18     1.82  40M     1:1
19     0.68  40M     1:5
20     0.83  40M     1:5
21     0.89  40M     1:5
22     0.58  40M    1:10
23     1.13  40M    1:10
24     1.28  40M    1:10
25     6.61 100M     1:0
```

```
26      6.66 100M     1:0
27      7.43 100M     1:0
28      1.25 100M     1:1
29      1.46 100M     1:1
30      1.49 100M     1:1
31      1.17 100M     1:5
32      1.27 100M     1:5
33      1.16 100M     1:5
34      0.93 100M    1:10
35      0.67 100M    1:10
36      0.80 100M    1:10
> interaction.plot(trifluralin$conc, trifluralin$delrat, trifluralin$sorption)
> interaction.plot(trifluralin$delrat, trifluralin$conc, trifluralin$sorption)
> boxplot(split(trifluralin$sorption, trifluralin$delrat:trifluralin$conc),
ylab="Sorption", xlab="DeliveryRatio:Concentration")
> title(main="Boxplot of sorption under all experimental conditions")
> boxplot(split(trifluralin$sorption, trifluralin$delrat:trifluralin$conc),
ylab="Sorption", xlab="DeliveryRatio:Concentration",log="y")
> title(main="Boxplot of sorption under all experimental conditions, log scale")
```




mean of trifluralin$sorption
2
4
6
8
10
trifluralin$conc
100M
40M
15M
1:0
1:1
1:10
1:5
trifluralin$delrat

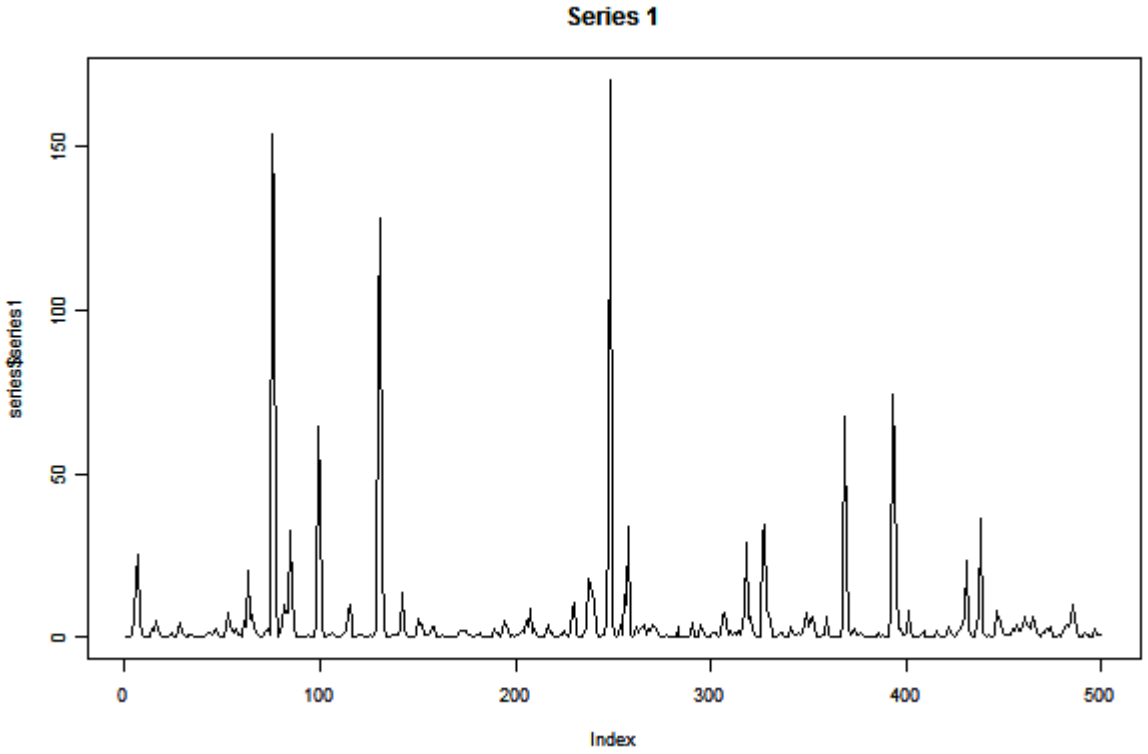

Boxplot of sorption under all experimental conditions
Sorption
2
4
6
8
10
12
1:0:100M
1:0:40M
1:1:15M
1:10:100M
1:10:40M
1:5:15M
DeliveryRatio:Concentration

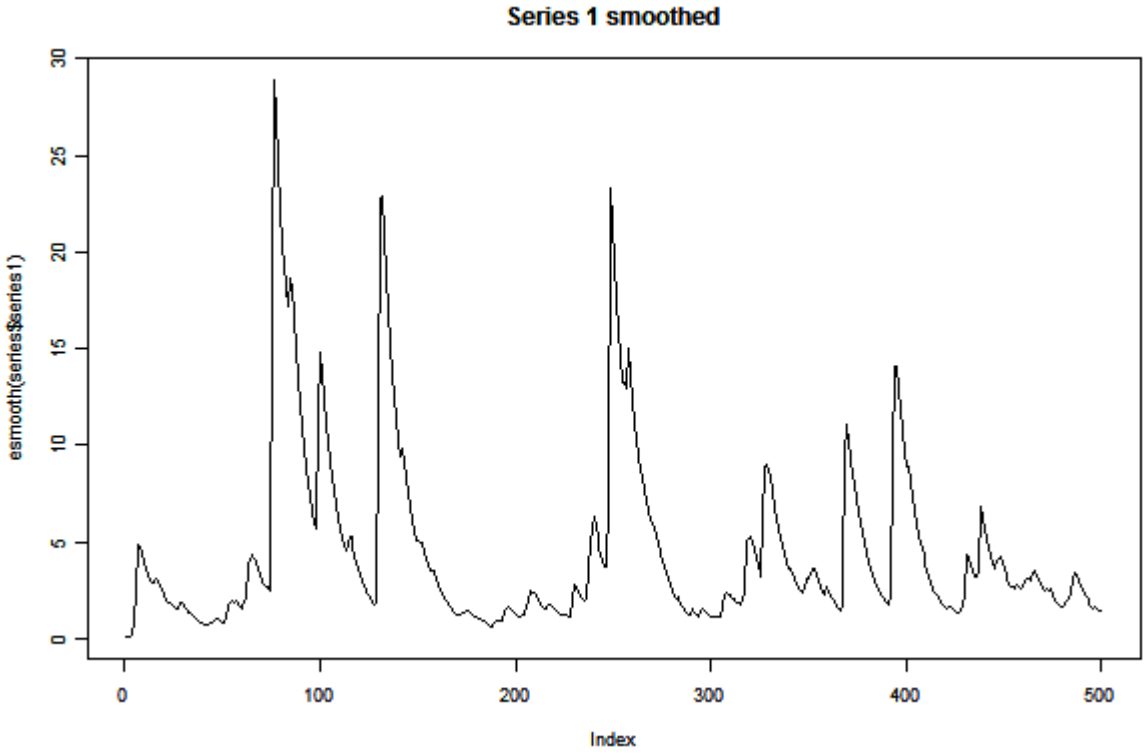

Boxplot of sorption under all experimental conditions, log scale
Sorption
0.5
1.0
2.0
5.0
10.0
1:0:100M
1:0:40M
1:1:15M
1:10:100M
1:10:40M
1:5:15M
DeliveryRatio:Concentration

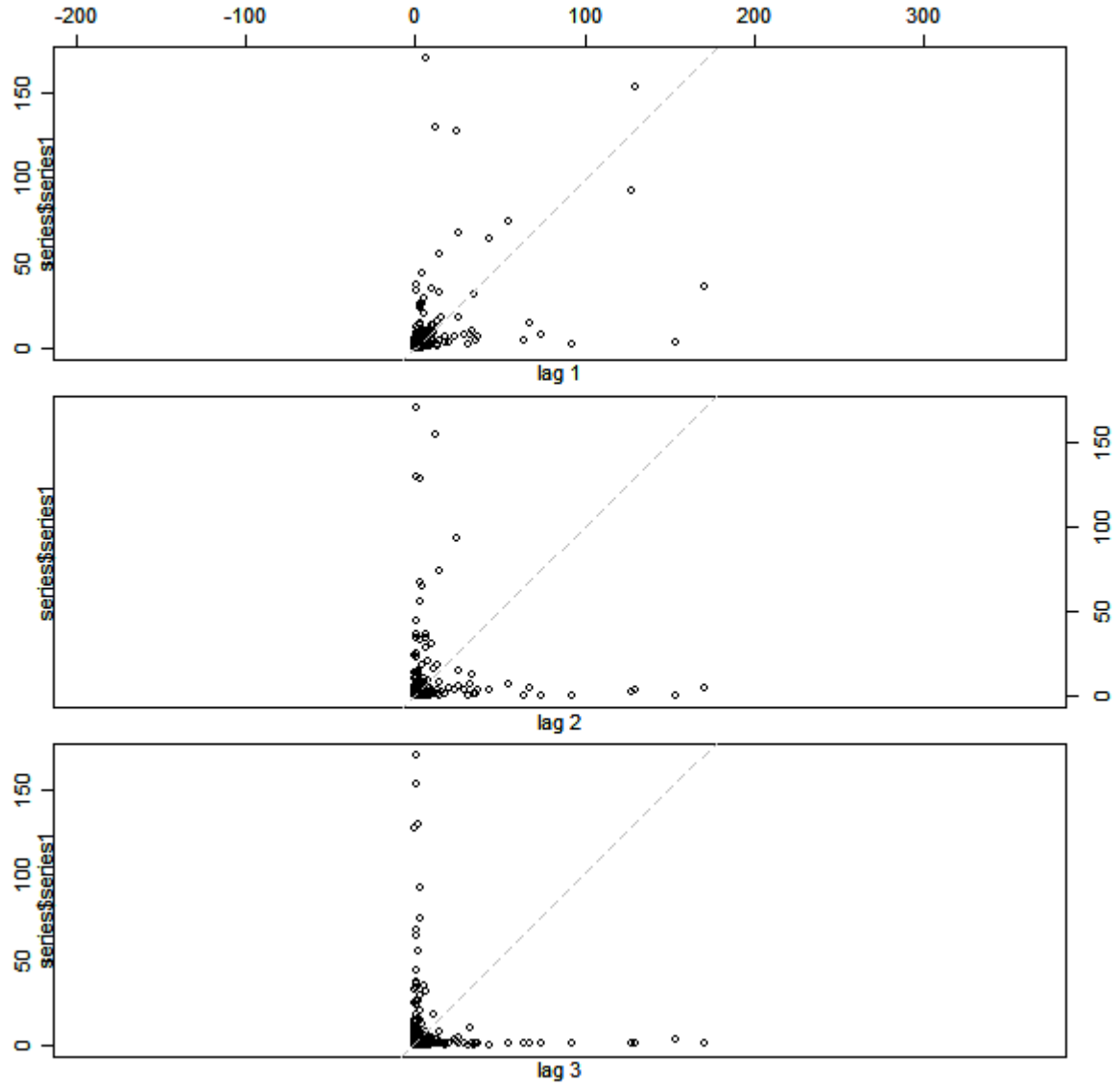
Question 3. **[12 marks for appropriate conclusions supported by graphs.]**

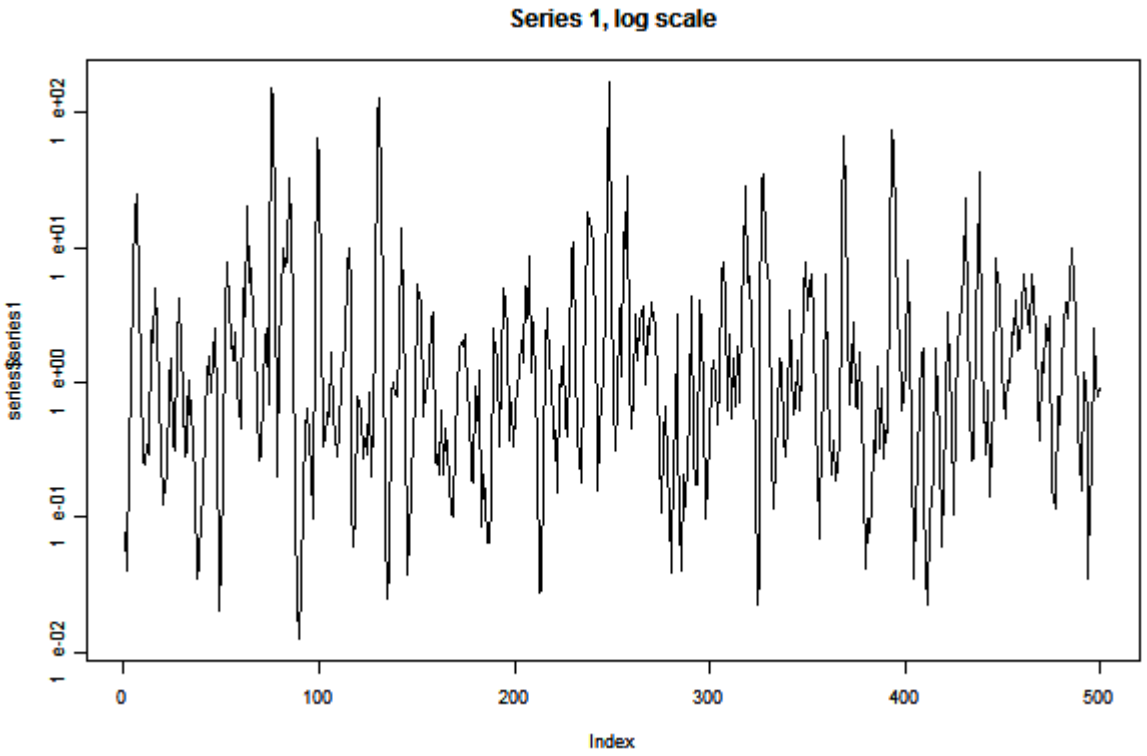
The lag plots clearly show that the observations in Series 2 are independent and observations in Series 3 are autocorrelated with lag 1. The lag plots for Series 1 are difficult to interpret on the original scale, but on the log scale they show autocorrelation with lag 1. The time sequence plots confirm this but the effect is less obvious; Series 2 oscillates on a scale of 1 time step, Series 3 oscillations are about 2 time steps wide. Series 1 appears to be very erratic with occasional spikes on the original scale but on the log scale it looks like Series 3. Smoothing the series does not seem to be helpful here for revealing autocorrelation. Note that with Series 1 it makes a difference whether we smooth before or after the log transformation.

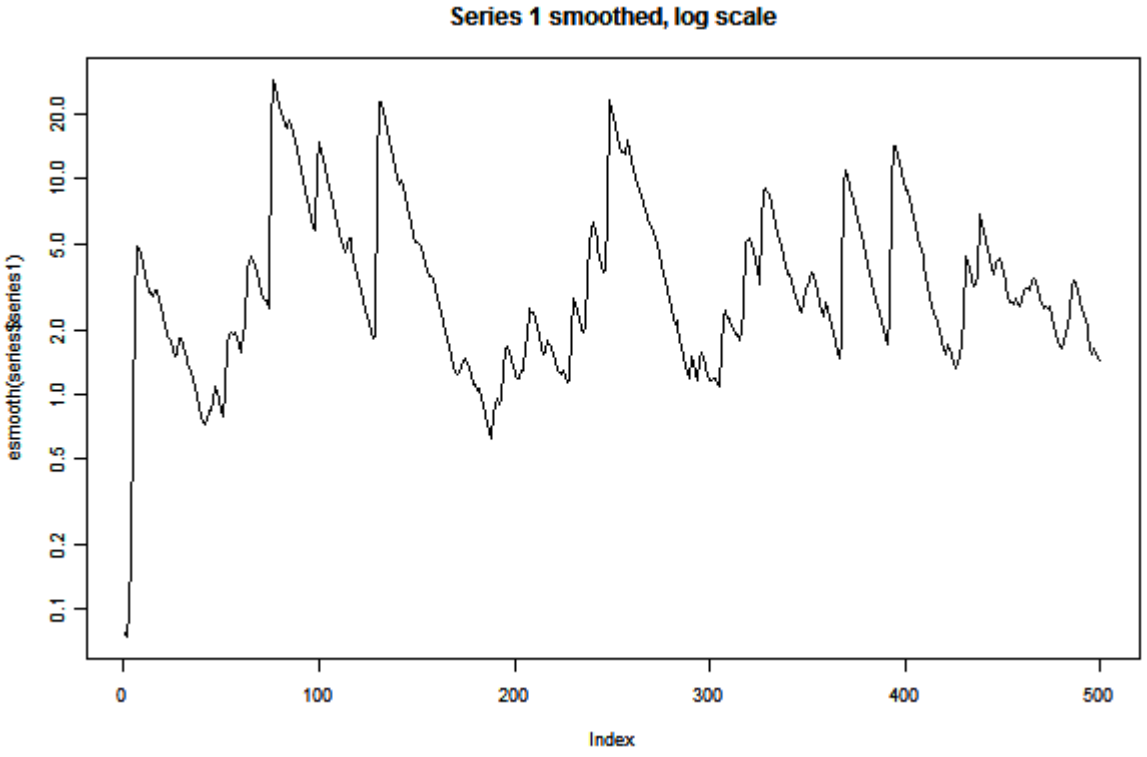
```
> smooth <- function (series, alpha = 0.1)
{
    esseries = series
    for (t in 2:length(series)) esseries[t] <- ifelse(is.na(series[t]),
        esseries[t - 1], alpha * series[t] + (1 - alpha) * esseries[t -
            1])
    esseries
}
> series <- read.table("series_b.txt")
> names(series)
[1] "series1" "series2" "series3"
> dim(series)
[1] 500   3
> plot(series$series1, type="l", main="Series 1")
> plot(esmooth(series$series1), type="l", main="Series 1 smoothed")
> lag.plot(series$series1, 3)
> plot(series$series1, type="l", log="y", main="Series 1, log scale")
> plot(esmooth(series$series1), type="l", log="y", main="Series 1 smoothed, log scale")
> plot(esmooth(log(series$series1)), type="l", main="Series 1, log scale, smoothed")
> lag.plot(log(series$series1), 3)
> plot(series$series2, type="l", main="Series 2")
> plot(esmooth(series$series2), type="l", main="Series 2 smoothed")
> lag.plot(series$series2, 3)
> plot(series$series3, type="l", main="Series 3")
> plot(esmooth(series$series3), type="l", main="Series 3 smoothed")
> lag.plot(series$series3, 3)
```

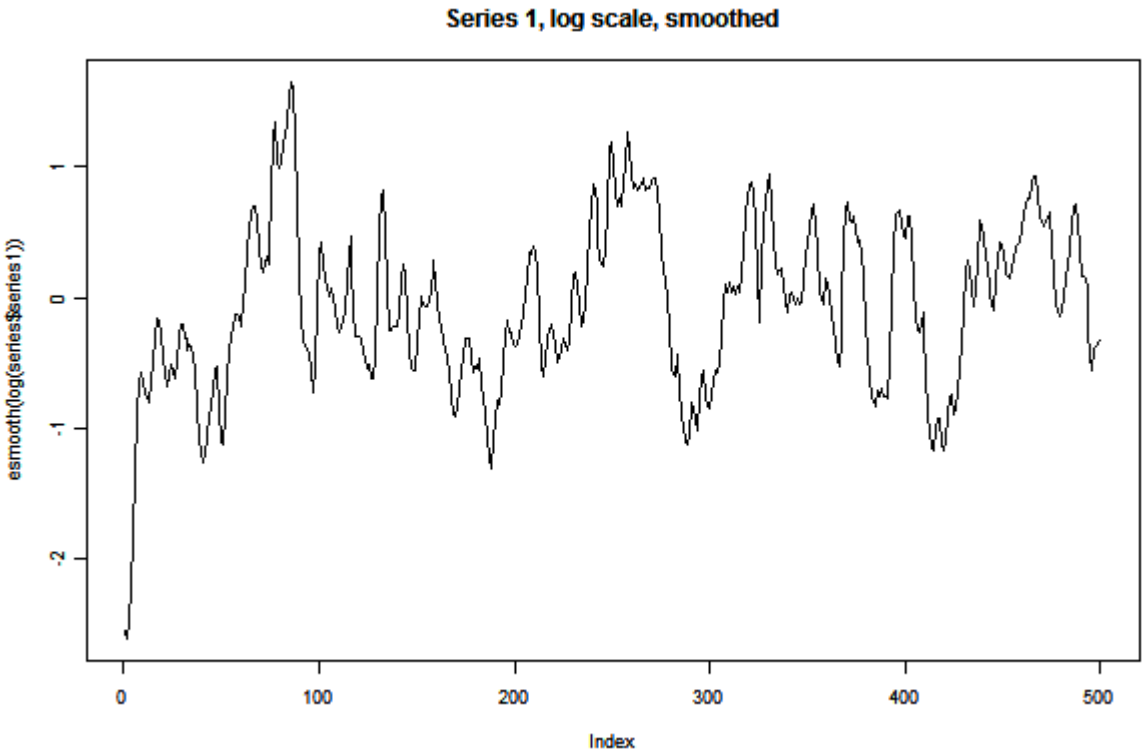

Series 1
series$series1
0
50
100
150
0
100
200
300
400
500
Index

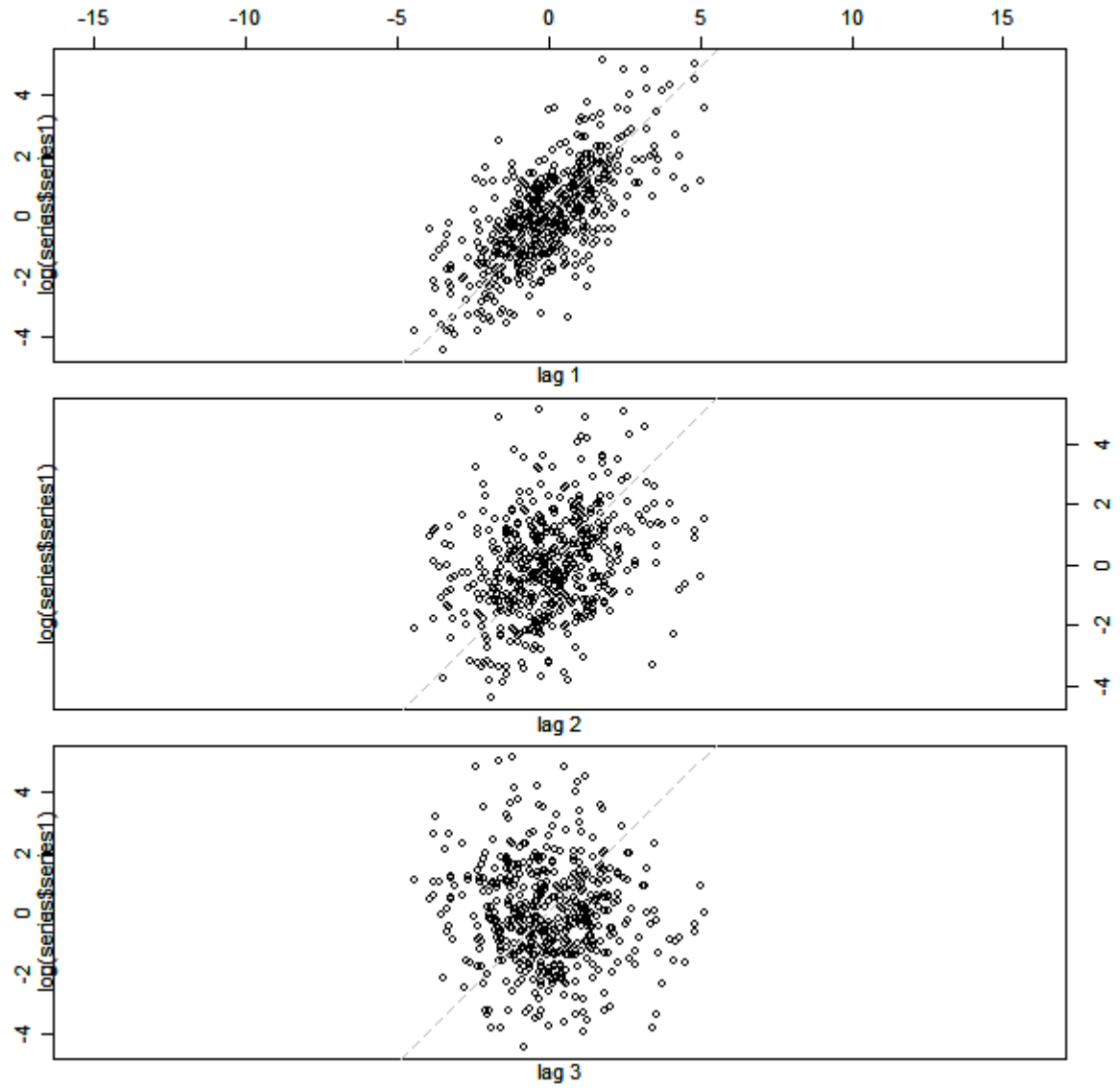

Series 1 smoothed
esmooth(series$series1)
0
5
10
15
20
25
30
0
100
200
300
400
500
Index

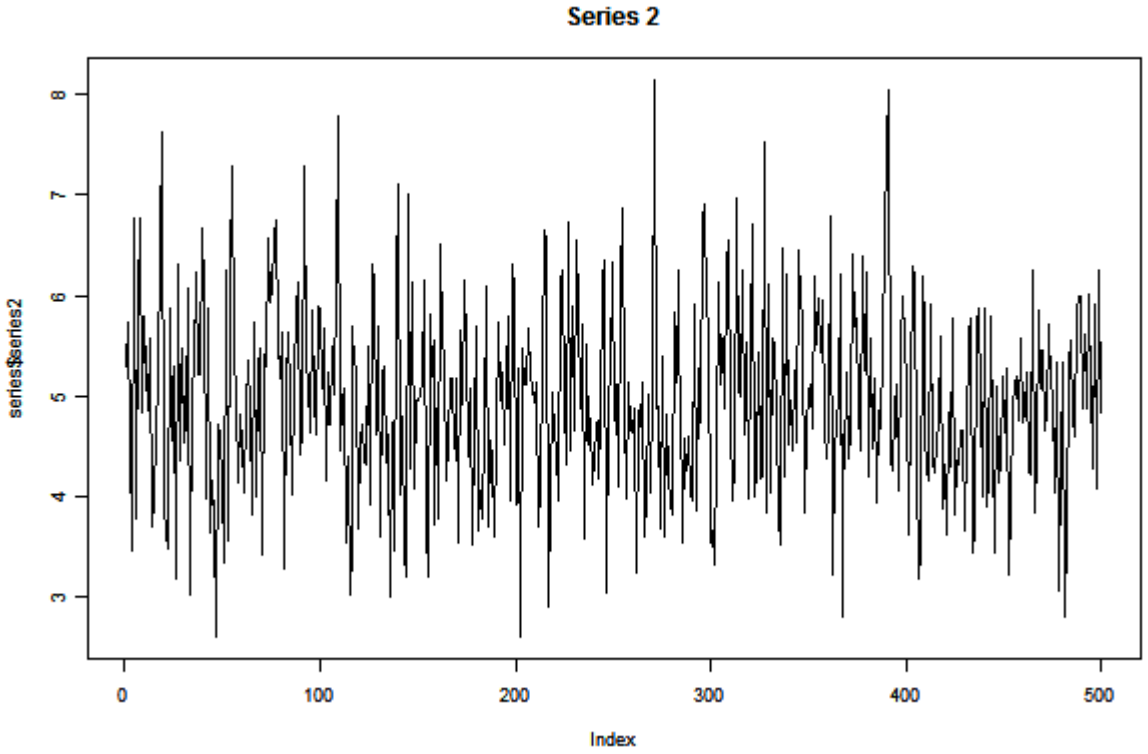

-200
-100
0
100
200
300
0
50
100
150
series$series1
lag 1
lag 2
lag 3

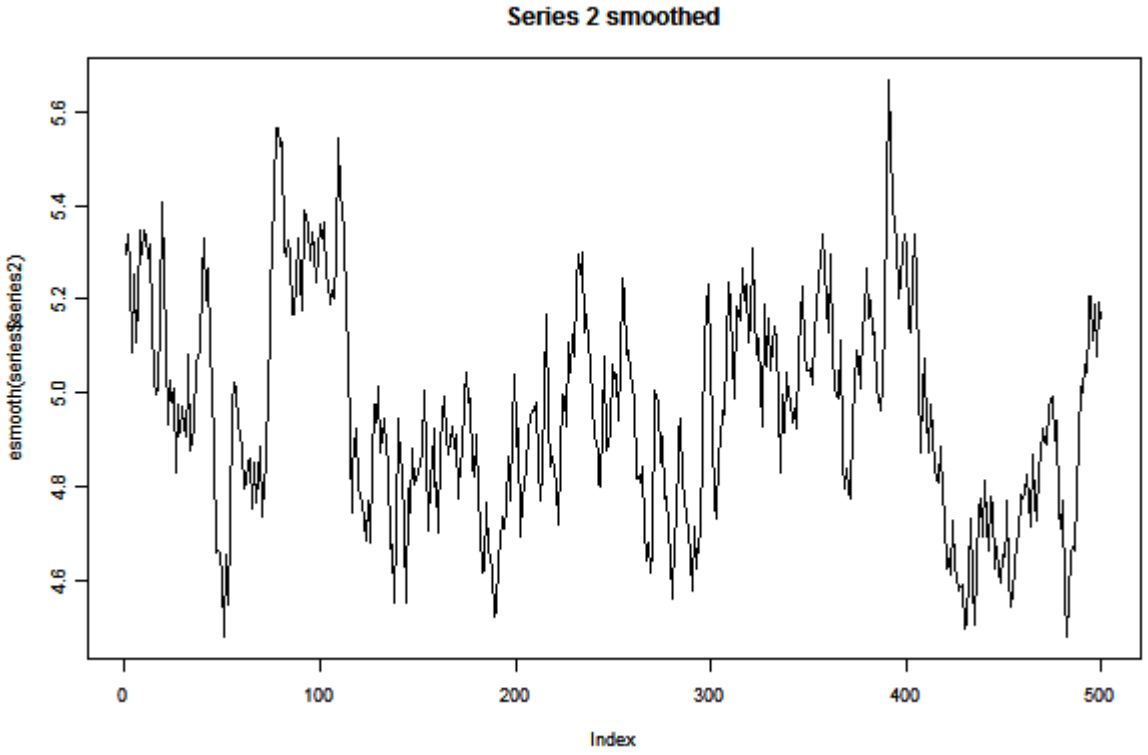

Series 1, log scale
series$series1
1 e-02
1 e-01
1 e+00
1 e+01
1 e+02
0
100
200
300
400
500
Index

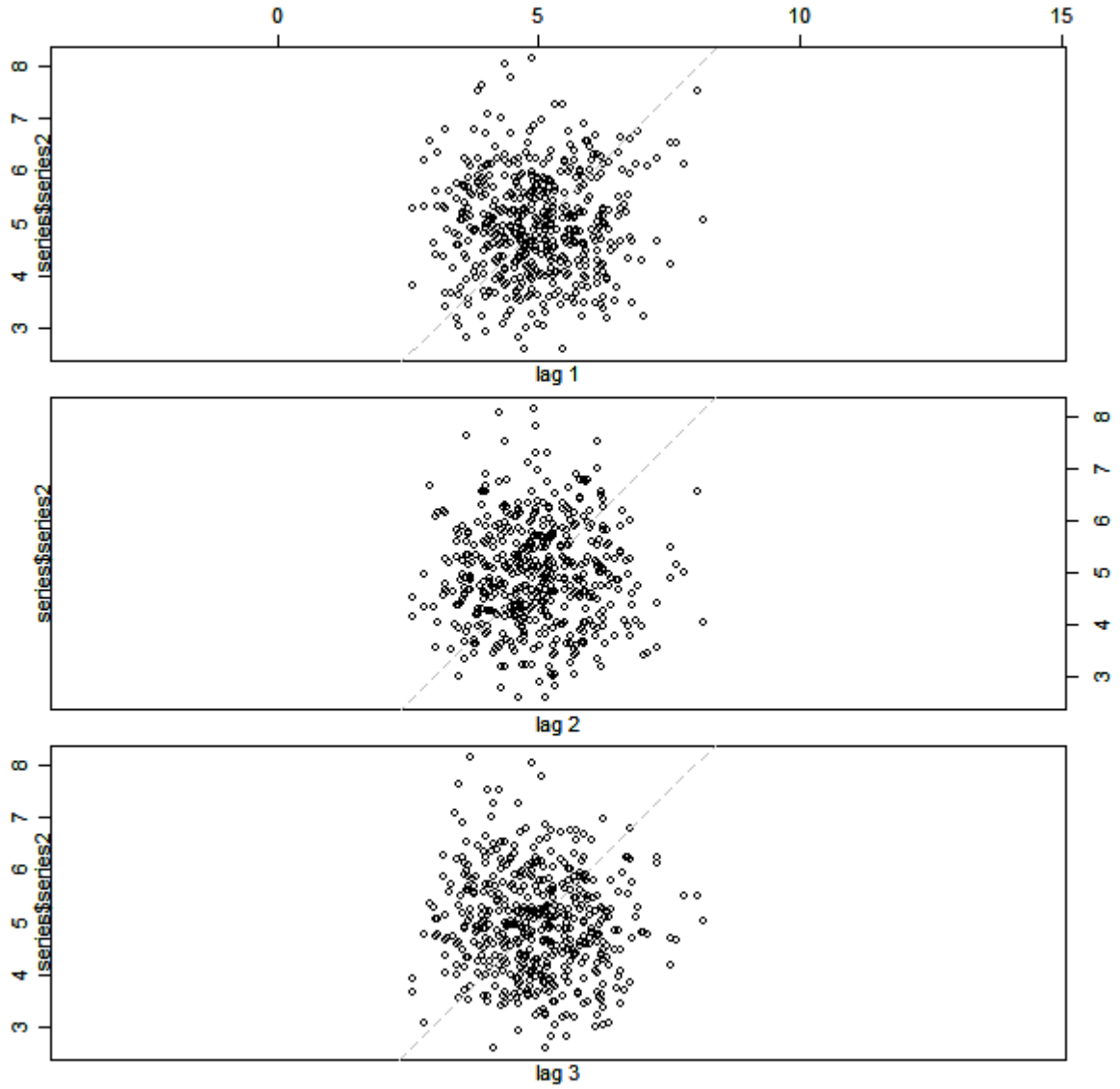

Series 1 smoothed, log scale
esmooth(series$series1)
0.1
0.2
0.5
1.0
2.0
5.0
10.0
20.0
0
100
200
300
400
500
Index

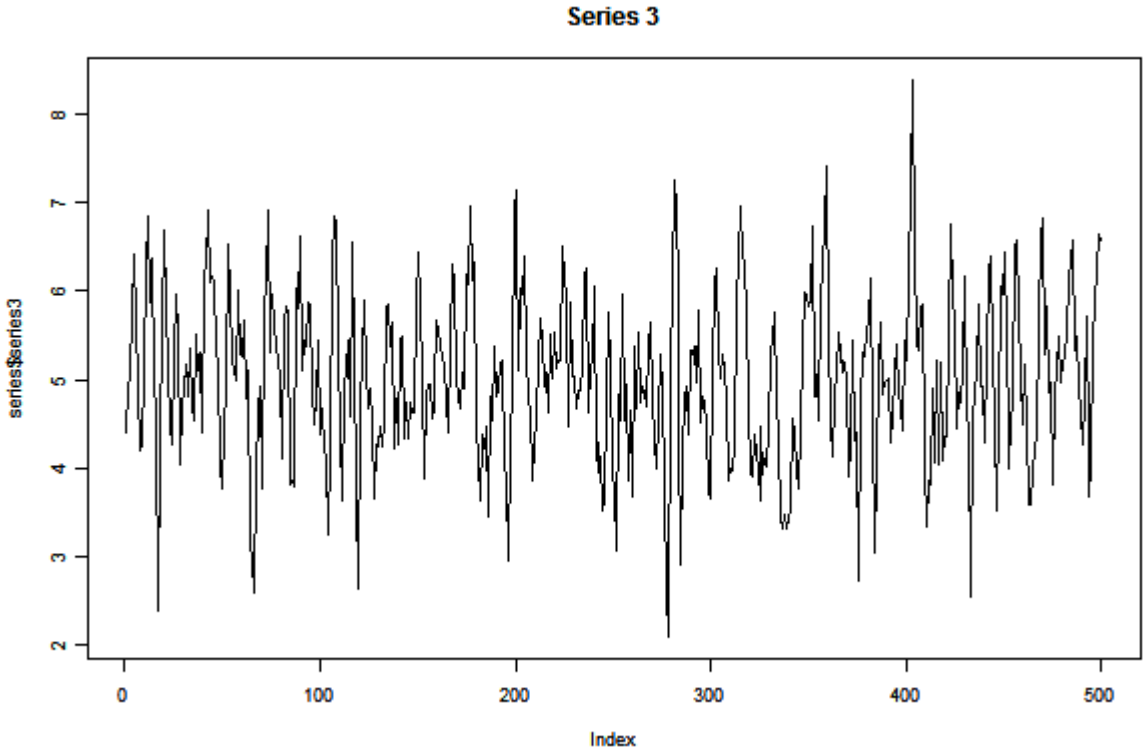

Series 1, log scale, smoothed
esmooth(log(series$series1))
1
0
-1
-2
0
100
200
300
400
500
Index

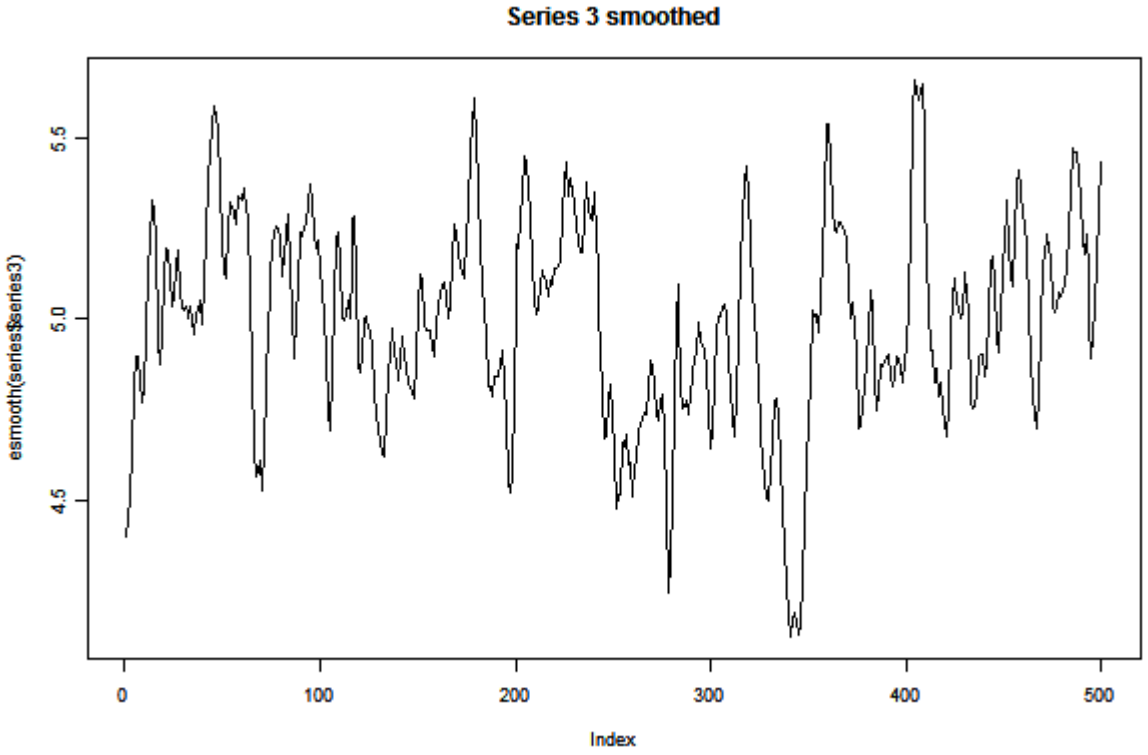

-15
-10
-5
0
5
10
15
log(seriesftseries1)
lag 1
lag 2
lag 3

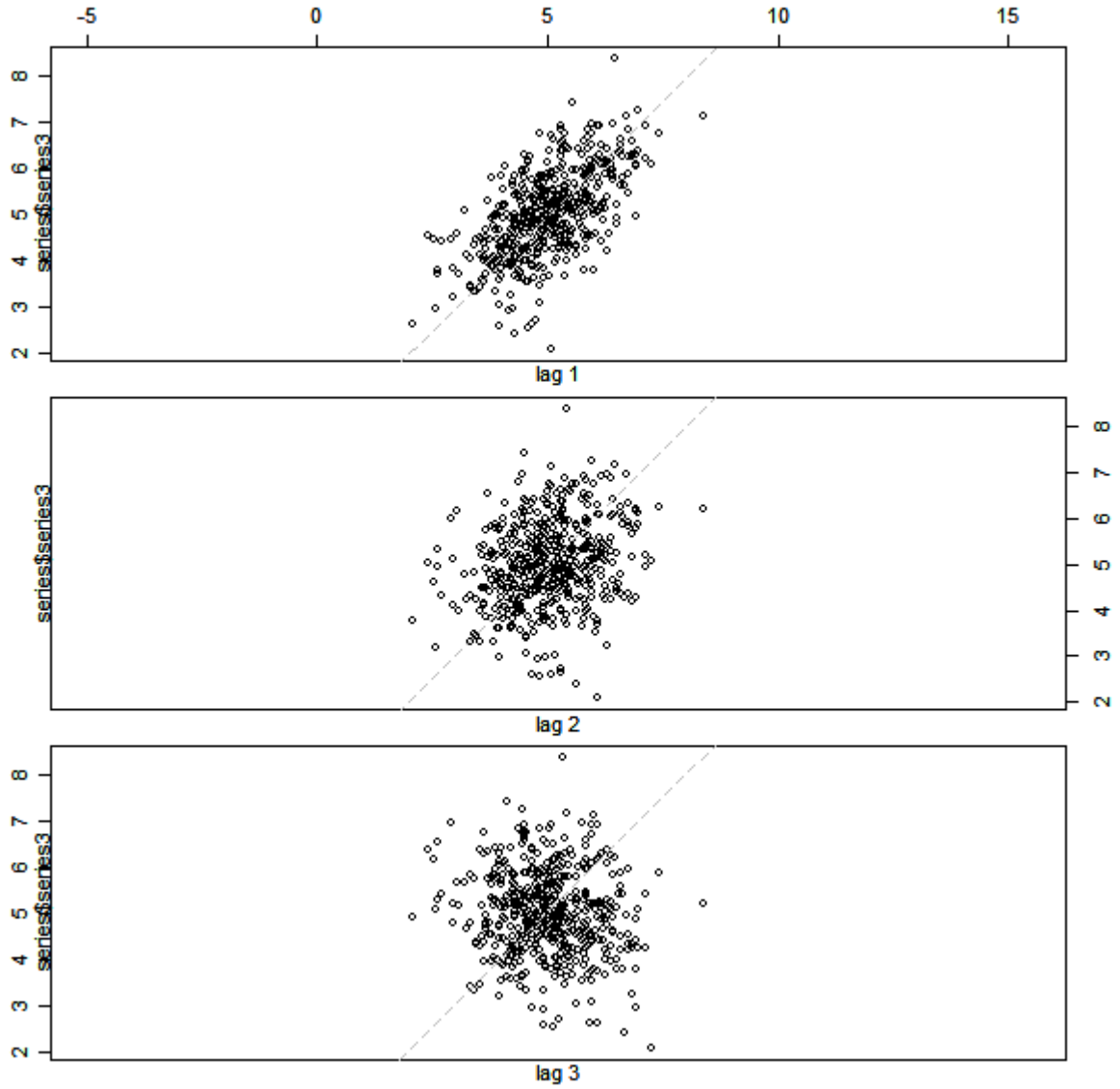

Series 2
series$series2
Index
0
100
200
300
400
500
3
4
5
6
7
8


Series 2 smoothed
esmooth(series$series2)
Index
0
100
200
300
400
500
4.6
4.8
5.0
5.2
5.4
5.6


0
5
10
15
series$series2
lag 1
lag 2
lag 3


Series 3
series3$series3
Index
0
100
200
300
400
500
2
3
4
5
6
7
8


Series 3 smoothed
esmooth(series3$series3)
Index
0
100
200
300
400
500
4.5
5.0
5.5



BONUS MARKS: [Up to 9 marks for the quality of writing and presentation]

FULL MARKS = 45