AMS 316.01: Homework 2

Due on 09/24

Haipeng Xing

Harris Temuri

Problem 1

The following data show the coded sales of company X in successive 4-week periods over 1995-1998.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
1995	153	189	221	215	302	223	201	173	121	106	86	87	108
1996	133	177	241	228	283	255	238	164	128	108	87	74	95
1997	145	200	187	201	292	220	233	172	119	81	65	76	74
1998	111	170	243	178	248	202	163	139	120	96	95	53	94

a. Plot the data.

b. Assess the trend and seasonal effects.

Solution

Part A

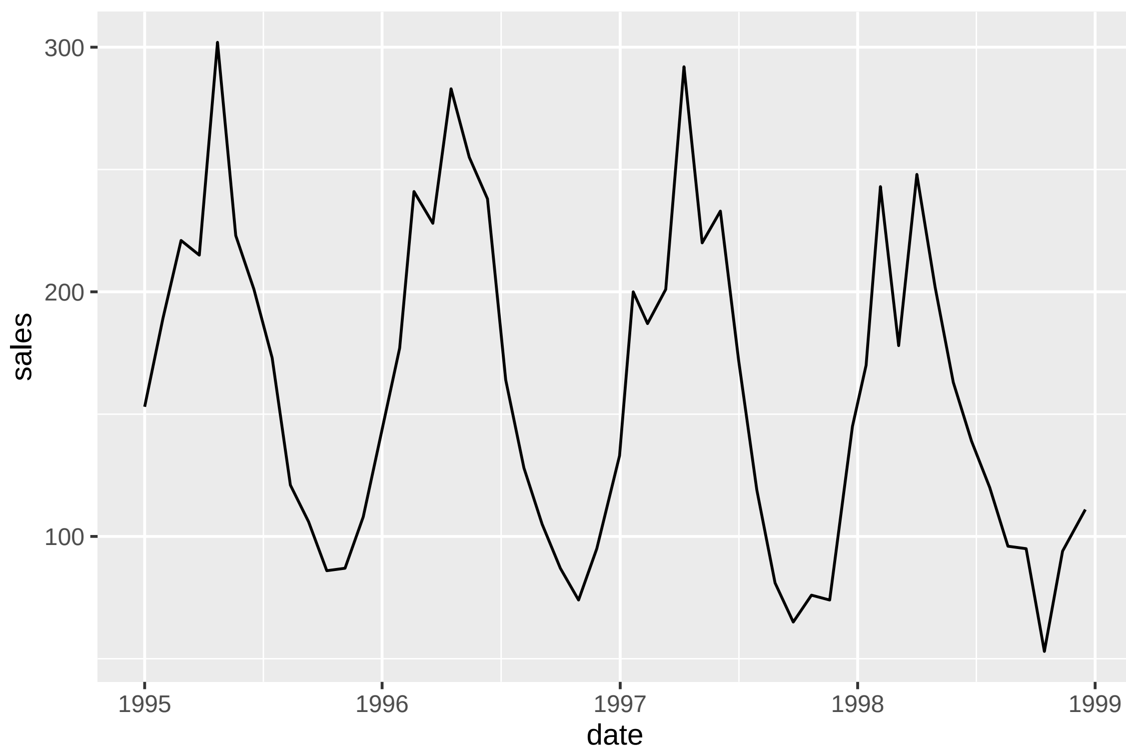
Create data on excel

	A	B	C
1	year	week(interval)	sales
2	1995	1	153
3	1995	2	189
4	1995	3	221
5	1995	4	215
6	1995	5	302
7	1995	6	223
8	1995	7	201
9	1995	8	173
10	1995	9	121
11	1995	10	106
12	1995	11	86
13	1995	12	87
14	1995	13	108
15	1996	1	133
16	1996	2	177
17	1996	3	241
18	1996	4	228
19	1996	5	283
20	1996	6	255
21	1996	7	238
22	1996	8	164
23	1996	9	128
24	1996	10	105
25	1996	11	87
26	1996	12	74
27	1996	13	95
28	1997	1	145
29	1997	2	200
30	1997	3	187
31	1997	4	201
32	1997	5	292
33	1997	6	220
34	1997	7	233
35	1997	8	172
36	1997	9	119

Import data to R-Studio

```
data_2 <- read_excel("C:/Users/Harris/Desktop/AMS 316/Homework 2/data_2.xlsx")
data_2$date <- as.Date('1995-01-01')
data_2$date <- as.Date(data_2$date, '%Y/%m/%d')
for (i in 1:51) {
 week(data_2$date[i+1]) <- week(data_2$date[i])+4
 year(data_2$date[i+1]) <- year(data_2$date[i])+(data_2$year-1995)
}
for (i in 1:51) {
 year(data_2$date[i+1]) <- year(data_2$date[i])+(data_2$year[i+1]-data_2$year[i])
}
ggplot(data = data_2, aes(date, sales)) + geom_line()
```

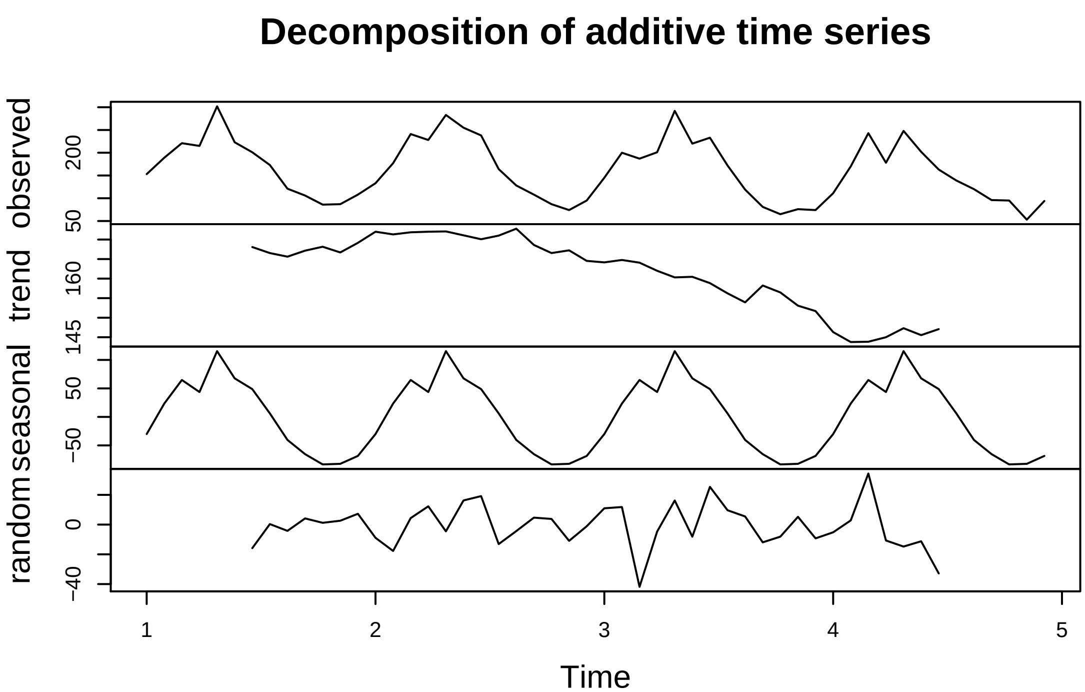
Figure 1: Plot of sales over time



Part B

```
x <- c(153, 189, 221, 215, 302, 223, 201, 173, 121, 106, 86, 87, 108, 133, 177, 241,
    228, 283, 255, 238, 164, 128, 108, 87, 74, 95, 145, 200, 187, 201, 292, 220, 233,
    172, 119, 81, 65, 76, 74, 111, 170, 243, 178, 248, 202, 163, 139, 120, 96, 95, 53,
    94)
ts_data <- ts(x, frequency = 13)
plot(decompose(ts_data))
```

Figure 2: Decomposed time series of Company X data



Looking at the graph, we can see a general downtrend and an additive seasonality.

Problem 2

Sixteen successive observations on a stationary time series are as follows: 1.6, 0.8, 1.2, 0.5,0.9, 1.1, 1.1, 0.6, 1.5, 0.8, 0.9, 1.2, 0.5, 1.3, 0.8, 1.2

a. Plot the data.

b. Looking at the graph plotted in (a), guess an approximate value for the auto-correlation coefficient at lag 1.

c. Plot x_t against X_{t+1}, and again try to guess the value of r_1

d. Calculate r_1

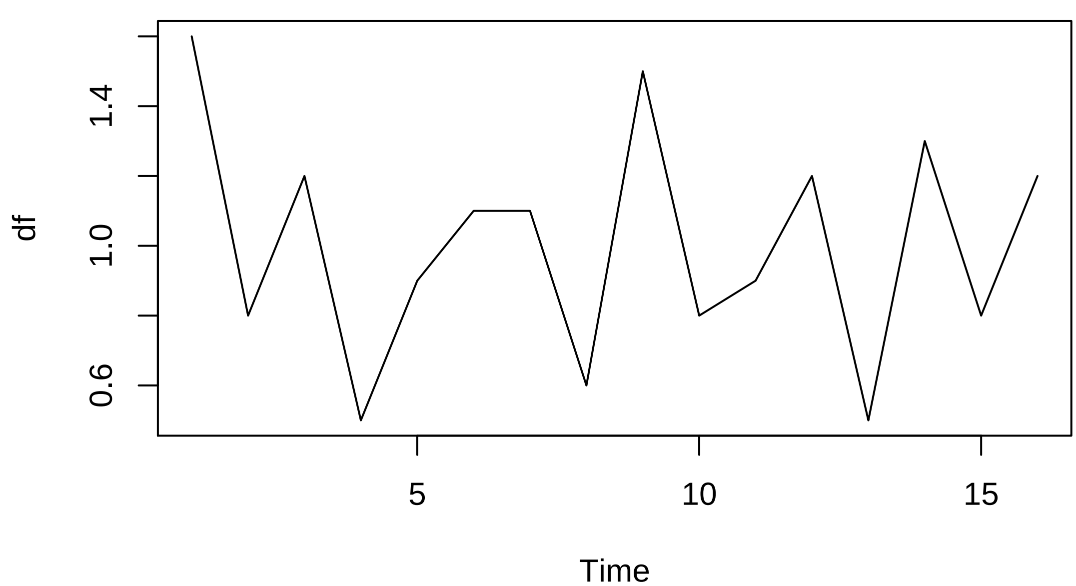
Solution

Part A

Create data frame

```
df <- c(1.6, 0.8, 1.2, 0.5,0.9, 1.1, 1.1, 0.6,1.5, 0.8, 0.9, 1.2, 0.5, 1.3, 0.8, 1.2)
plot.ts(df)
plot(df)
```

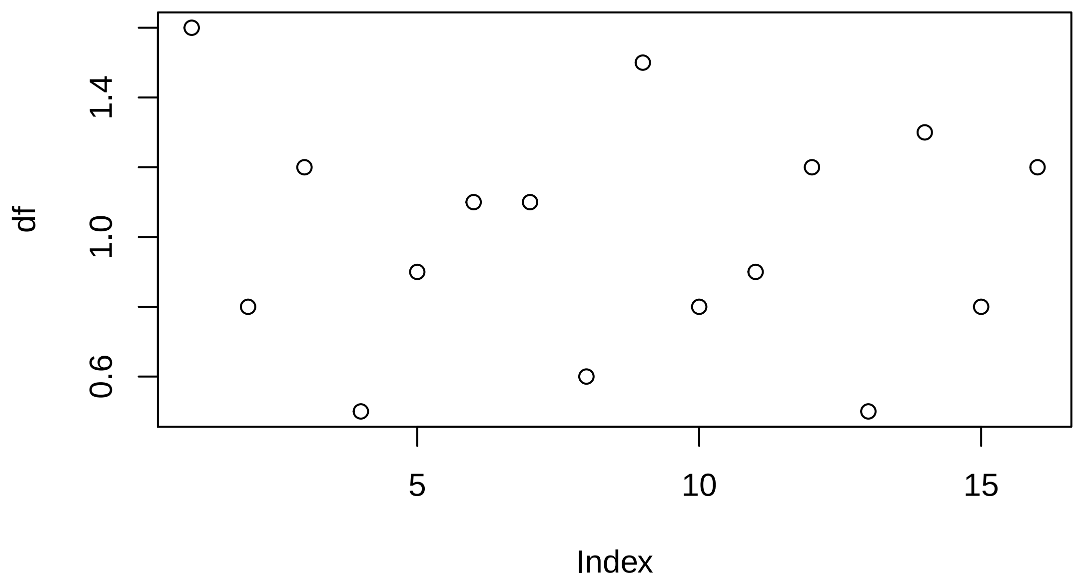
Figure 3: Line plot of observations over time



Problem 2 continued on next page...

```
plot(df)
```

Figure 4: Scatter plot of observations over time



Part B

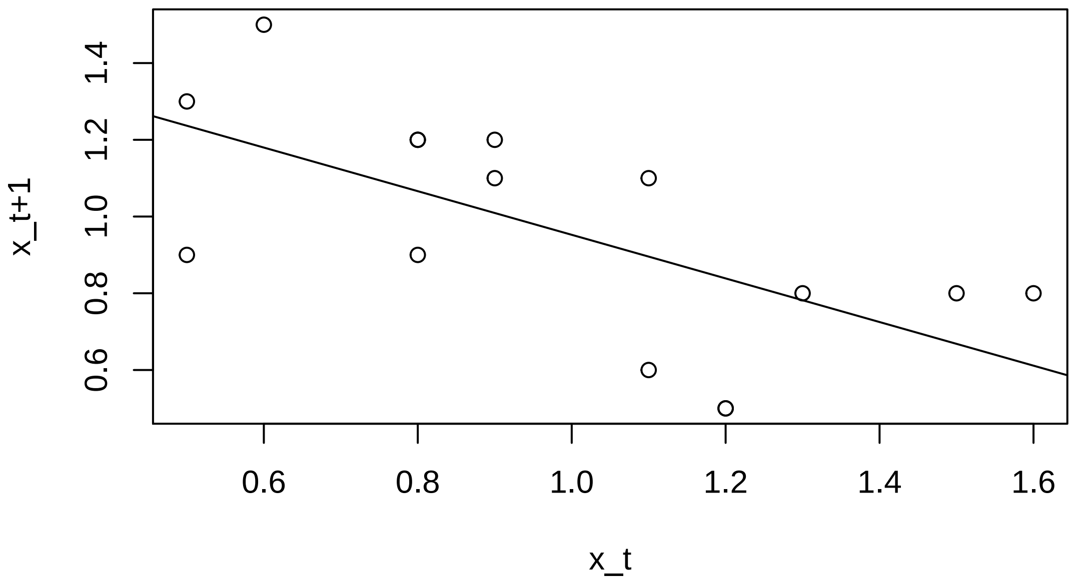
Looking at the graph, I cannot really tell what the value of r_1 might be.

Problem 2 continued on next page. . .

Part C

```
x_t <- df[1:15]
'x_t+1' <- df[2:16]
plot(x_t, 'x_t+1')
abline(lm('x_t+1' ~ 'x_t))
```

Figure 5: Scatter plot with trend line of x_t vs x_{t+1}



Looking at this plot, I would say r_1 is about -0.6.

Part D

```
acf(df)[]

# Autocorrelations of series 'df', by lag
#
#     0      1      2      3      4      5      6      7      8      9     10
# 1.000 -0.549  0.250 -0.104 -0.165  0.067  0.037 -0.116  0.122  0.098 -0.189
#    11     12
# 0.220 -0.305
```

The value of r_1 is -0.549.

Problem 3

A computer generates a series of 400 observations that are supposed to be random. The first 10 sample auto-correlation of the series are $r_1 = 0.02$, $r_2 = 0.05$, $r_3 = -0.09$, $r_4 = 0.08$, $r_5 = -0.02$, $r_6 = 0.00$, $r_7 = 0.12$, $r_8 = 0.06$, $r_9 = 0.02$, $r_{10} = -0.08$. Is there any evidence of non-randomness?

Solution

According to your slides,

> ...if a time series is random, we can expect 19 out of 20 of the values of r_k to lie between $\pm 1.96/\sqrt{N}$.

$$\pm 1.96/\sqrt{400} = \pm 0.98$$

Looking at the values of the auto-correlation, it seems there is one value which constitutes as significant, $r_7 = 0.12$. However since there is only value out of the 10 which is significant, I don't believe it constitutes as evidence of non-randomness.

Problem 4

Suppose we have a seasonal series of monthly observation $\{X_t\}$, for which the seasonal factor at time t is denoted by $\{S_t\}$. Further suppose that seasonal pattern is constant through time so that $S_t = S_{t-12}$ for all t. Denote a stationary series of random deviations by $\{e_t\}$.

a. Consider the model $X_t = a + b_t + S_t + e_t$ having a global linear trend and multiplicative seasonality. Show that the seasonal difference operator ∇_{12} acts on X_t to produce a stationary series.

b. Consider the model $X_t = (a + b_t)S_t + e_t$ having a global linear trend and multiplicative seasonality. Does the operator ∇_{12} transform X_t to stationary? If not, define a difference operator that does. (Note: As stationary is not formally defined until Chapter 3, you should use heuristic arguments. A stationary process may involve a constant mean value (that could be zero) plus any linear combination of the stationary series $\{e_t\}$, but should not include terms such as trend and seasonality.)

Solution

Part A

Since the data has a non varying seasonality, in order to transform the data into a stationary series, we need to remove the trend and seasonality. Using seasonal differencing(first or second order) will remove the both of them for us.

$$
\begin{aligned}
X_t &= a + b_t + S_t + e_t \\
\nabla_{12} &= X_t - X_{t-12} \\
&= a + b_t + S_t + e_t - a - b_{t-12} - S_{t-12} - e_{t_12} \\
&= (b_t - b_{t-12}) + (S_t - S_{t-12}) + e_t - e_{t-12} \\
&= c + 0 + e_t - e_{t-12}; c \text{ is a constant}
\end{aligned}
$$

The first order seasonal differencing removed both the trend and seasonality from this model.

Part B

$$X_t = (a + b_t)S_t + e_t$$
$$\nabla_{12} = x_t - x_{t-12}$$
$$= aS_t + b_tS_t + e_t - aS_{t-12} - b_{t-12}S_{t-12} - e_{t-12}$$
$$= 0 + (b_tS_t - b_{t-12}S_{t-12}) + (e_t - e_{t-12})$$
$$= cS_t + e_t - e_{t-12}; c \text{ is a constant}$$

Still has seasonality. Will try second order seasonal differencing.

$$X_t = (a + b_t)S_t + e_t$$
$$\nabla^2_{12} = x_t - 2x_{t-12} + x_{t-11}$$
$$= (a + b_t)S_t + e_t - 2((a + b_{t-12})S_{t-12} + e_{t_12}) + (a + b_{t-11})S_{t-11} + e_{t-11}$$
$$= (a_{St} - 2aS_t + aS_t) + (b_tS_t - 2b_{t-12}S_t + b_{t-11}S_t) + (e_t - 2e_{t-12} + e_{t-11})$$
$$= 0 + (b_t - 2b_{t-12} + b_{t-11})S_t + (e_t - 2e_{t-12} + e_{t-11})$$
$$= (b_t - b_t)S_t + (e_t - 2e_{t-12} + e_{t-11})$$
$$= e_t - 2e_{t-12} + e_{t-11}$$

The second order seasonal differencing removed both the trend and seasonality from this model.