

資料科學概論期末報告

從環境資源資料庫中可得近年溫度資訊，本資料使用西元 2018 年 1 月 1 日至 2019 年 1 月 2 日，台北市萬華空氣品質監測站監測資料進行分析及預測後五天氣溫走勢，使用軟體為 R。(2018 年 8 月 17 日為缺值，故將此日期拿掉，以便進行資料分析)

首先，先將得到的檔案進行整理，因原檔案時間是從 2019 年 1 月 2 日至 2018 年 1 月 1 日，所以須將日期反過來，變成從 2018 年 1 月 1 日至 2019 年 1 月 2 日。

程式：

```
> setwd("C://test")
> data0<-read.table("data0.csv",sep=";",header=F)
> Z<-data0[,1]
> length(Z)
[1] 364
> ZZ<-Z[seq(364,1,by=-1)]
> ZZ
[1] 17.95 20.88 22.22 21.41 17.85 17.89 18.89 18.82 12.56 12.65 12.70 11.76
[13] 13.49 17.18 19.78 19.64 20.87 21.40 19.60 19.73 20.29 18.50 19.36 17.71
[25] 17.77 16.60 17.95 17.00 13.38 13.75 15.95 12.80 13.81 11.07 9.99 8.95
[37] 11.26 12.43 13.40 16.73 17.23 13.39 14.70 16.36 19.02 21.48 19.66 18.22
[49] 21.08 21.84 19.32 17.40 15.88 17.09 18.41 16.48 17.66 20.31 18.90 21.86
[61] 22.83 22.29 24.62 22.94 17.68 19.94 14.84 14.30 16.88 20.04 20.77 21.02
[73] 23.09 22.73 21.03 20.28 23.18 24.56 20.02 16.15 15.78 19.00 21.90 23.27
[85] 23.43 22.34 24.04 25.63 24.30 24.53 25.41 25.21 25.28 25.51 26.06 20.50
[97] 16.28 20.14 22.44 23.68 25.95 25.95 27.41 26.90 20.02 20.10 18.19 20.97
[109] 25.00 25.62 26.82 27.88 28.35 23.80 21.48 23.40 24.33 25.04 26.55 27.23
[121] 28.00 29.22 22.84 24.77 27.11 28.23 29.33 25.90 23.63 23.83 26.44 27.59
[133] 28.73 29.12 29.16 29.80 30.72 30.96 29.95 30.27 29.35 29.75 27.54 27.77
[145] 29.83 30.85 30.95 30.78 30.48 30.22 30.30 25.20 25.49 27.44 28.77 30.03
[157] 29.11 29.94 30.53 29.04 28.30 26.57 28.68 28.44 26.73 25.65 27.10 28.94
[169] 29.32 29.99 30.65 30.05 29.19 29.90 30.13 30.74 31.00 31.40 31.69 29.84
[181] 30.06 30.57 32.28 32.08 32.38 30.80 31.18 29.34 30.38 31.33 28.91 29.28
[193] 30.81 31.66 31.51 30.98 30.51 31.73 31.65 31.11 30.34 29.33 29.77 30.39
[205] 30.04 30.69 31.48 30.80 31.02 32.05 32.44 32.88 32.53 32.73 30.56 30.08
[217] 31.50 31.50 31.43 31.19 30.76 31.15 31.07 29.81 30.64 28.86 30.04 30.86
[229] 29.83 29.20 29.52 30.05 30.96 30.40 28.60 29.66 29.33 28.78 28.20 28.22
[241] 27.47 27.86 28.11 30.49 31.61 30.38 30.55 29.42 27.98 25.56 25.72 30.41
```

[253] 30.26 31.16 30.45 29.83 30.90 32.16 31.65 31.05 30.40 29.99 29.73 29.35
 [265] 27.51 28.09 25.51 26.48 25.86 24.52 25.07 26.73 25.51 25.09 23.28 25.84
 [277] 26.19 24.53 25.02 25.38 24.50 22.37 20.75 22.08 24.48 25.53 24.98 22.28
 [289] 23.90 22.28 23.04 24.41 23.99 24.53 25.83 25.81 23.09 21.20 22.08 23.45
 [301] 21.03 20.31 22.35 24.48 24.94 24.68 26.05 26.93 25.74 23.45 24.55 25.36
 [313] 24.23 22.45 22.03 23.88 24.82 22.30 24.53 21.71 23.03 24.54 19.79 18.88
 [325] 22.23 24.23 22.00 21.53 20.55 21.14 23.02 23.55 24.98 23.89 24.66 22.64
 [337] 24.21 19.62 18.67 18.66 21.32 20.87 16.70 17.56 18.25 19.82 20.43 16.63
 [349] 17.80 20.72 21.11 22.46 23.43 21.31 17.96 21.57 20.98 18.31 15.60 14.34
 [361] 15.83 17.07 16.97 17.73

確認資料無誤後，開始進行分析。

程式：

```
> par(mfrow=c(2,2))
```

```
> ts.plot(Z)
```

```
> acf(Z)
```

```
> pacf(Z)
```

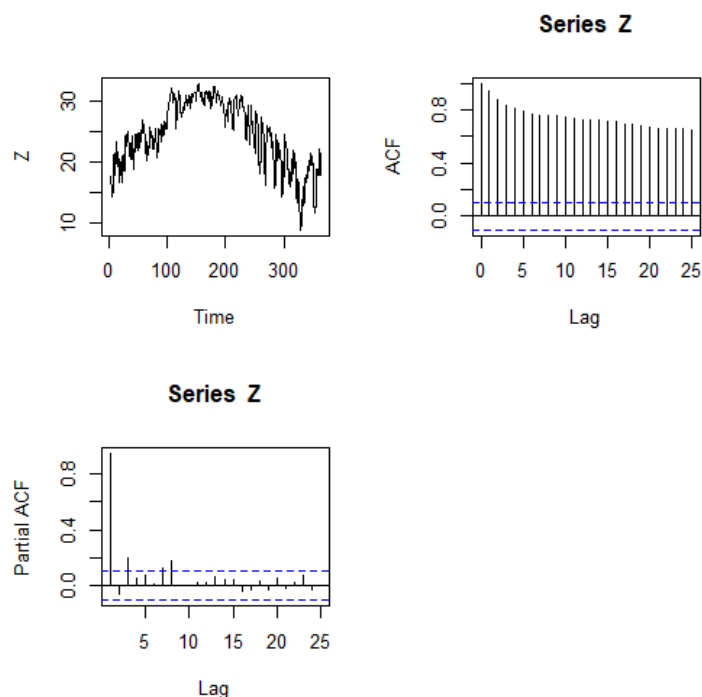
```
> adf.test(Z)
```

Error in adf.test(Z): 沒有這個函數 "adf.test"

從輸入的程式中可得圖(一)，而從圖(一)中可得此知數列不穩定，遇到不穩定的時間數列時，在

分析前進行差分。

圖(一)



程式：

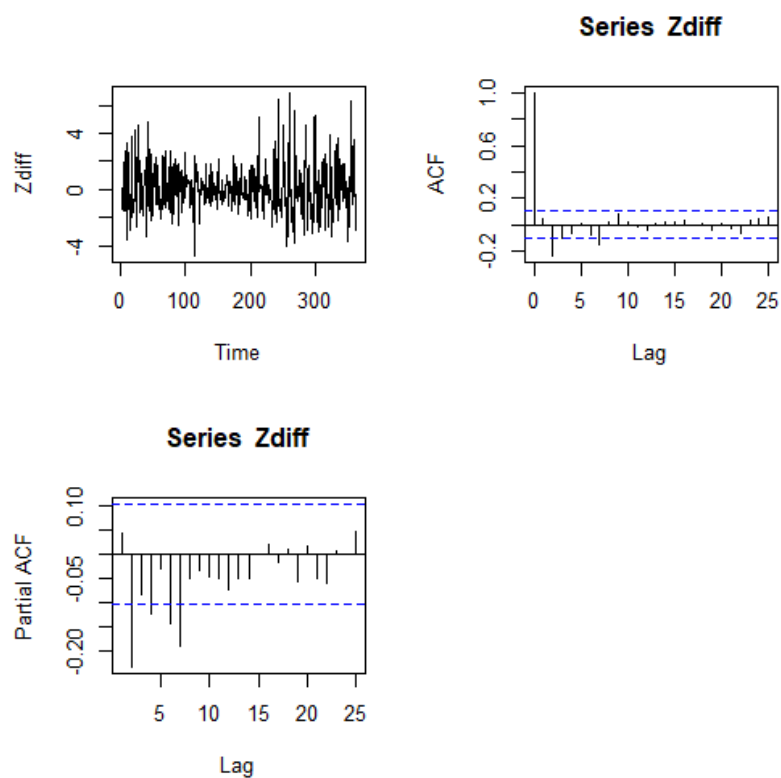
```
> Zdiff<-diff(Z)
```

```
> adf.test(Zdiff)
```

Error in adf.test(Zdiff)：沒有這個函數 "adf.test"

差分後可得圖(二)，進行一次差分後可見 ACF 圖形呈現快速下降趨勢，表示數列已穩定，得以進行分析。

圖(二)



程式：

```
> windows()
```

```
> par(mfrow=c(2,2))
```

```
> ts.plot(Zdiff)
```

```
> acf(Zdiff)
```

```
> pacf(Zdiff)
```

```
> for (p in 0:3)
```

```
+ {
```

```
+ for (q in 0:3)
```

```
+ {
```

```
+ aic_res<-arima(Zdiff,order=c(p,0,q))$aic
```

```
+ aic_res<-round(aic_res,3)
```

```
+ res<-paste("p=",p,",q=",q,",aic=",aic_res,sep="")
```

```

+ print(res)
+ }
+ }
[1] "p=0,q=0,aic=1463.789"
[1] "p=0,q=1,aic=1464.432"
[1] "p=0,q=2,aic=1439.961"
[1] "p=0,q=3,aic=1431.104"
[1] "p=1,q=0,aic=1465.048"
[1] "p=1,q=1,aic=1459.496"
[1] "p=1,q=2,aic=1423.065"
[1] "p=1,q=3,aic=1424.989"
[1] "p=2,q=0,aic=1446.486"
[1] "p=2,q=1,aic=1425.215"
[1] "p=2,q=2,aic=1424.931"
[1] "p=2,q=3,aic=1426.702"
[1] "p=3,q=0,aic=1445.869"
[1] "p=3,q=1,aic=1425.995"
[1] "p=3,q=2,aic=1426.463"
[1] "p=3,q=3,aic=1426.463"

```

Warning message:

In arima(Zdiff, order = c(p, 0, q)) :

possible convergence problem: optim gave code = 1

從上述結果可得知，當 $p=1$ ， $q=2$ 時，可得最小 $aic=1423.065$ ，將此數值帶入來預測後五天

氣溫。

程式：

```
> arima.res<-arima(Z,order=c(1,1,2))
```

```
> predict(arima.res,5)$pred
```

Time Series:

Start = 365

End = 369

Frequency = 1

```
[1] 17.39901 17.67791 17.82116 17.89473 17.93252
```

差分後數列符合 $ARMA(1,2)$ ，表示原數列為 $ARIMA(1,1,2)$ 。最後，可預測後五天氣溫為

17.39901、17.67791、17.82116、17.89473 及 17.93252。