IDS 702: Module 7.1

Introduction to time series analysis

Dr. Olanrewaju Michael Akande

Introduction

- When data are ordered in time, responses and errors from one period may influence responses and errors from another period.
- For example, it is reasonable to expect unemployment rate in a month to be correlated with unemployment rate in previous month(s).
- Another example: weather events in current time period may depend on weather events in previous time period.
- These are called time series data.
- Correlation due to time is called serial correlation or autocorrelation.
- We will only scratch the surface in this course.

Goals of time series analysis

- Forecasting outcomes
  - Given a series of outcomes ordered in time, predict the values of the outcomes in the future.
  - Examples:
    - forecasting future price of oil given historical oil prices.
    - predicting future price of a particular stock price given past prices of the same stock.
- When forecasting, it is important to also report an interval estimate to incorporate uncertainty about future values.

IDS 702

Goals of time series analysis

- Forecasting outcomes
    - Forecasting outcomes using predictors may involve building a model for the predictors as well, since we can't observe them in the future.
    - For example, predicting inflation rate given employment rate requires estimating future values for the employment rate as well.
- Learning relationships with data ordered in time.
    - How are outcomes correlated over time? Are there periodic relationships in outcomes?
    - Regressions of outcomes on predictors, accounting for correlated errors due to time series.

IDS 702

Motivating example: FTSE 100

- The FTSE (Financial Times Stock Exchange) 100 Index is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization.

- A share index is essentially a form of weighted average of prices of selected stocks.

- To motivate our discussions on time series, let's look at data for FTSE 100 returns in 2018.

```
ftse100 <- read.csv("data/ftse2018.csv", header = T)
head(ftse100)
```

```
##         Date    Open    High     Low   Close
## 1  11/7/2018 7040.68 7136.75 7040.68 7117.28
## 2  11/6/2018 7103.84 7117.50 7027.45 7040.68
## 3  11/5/2018 7094.12 7140.37 7077.40 7103.84
## 4  11/2/2018 7114.66 7196.39 7094.12 7094.12
## 5  11/1/2018 7128.10 7165.61 7085.74 7114.66
## 6 10/31/2018 7035.85 7161.54 7035.85 7128.10
```

- Can we forecast closing prices for the next five days from 11/7/2018?

IDS 702

Motivating example: FTSE 100

Notice that the data go from latest to earliest date, so let's invert the order of the rows to make the time series increasing in date.

```
ftse100 <- ftse100[nrow(ftse100):1,]
dim(ftse100)
```

```
## [1] 211   5
```

```
head(ftse100)
```

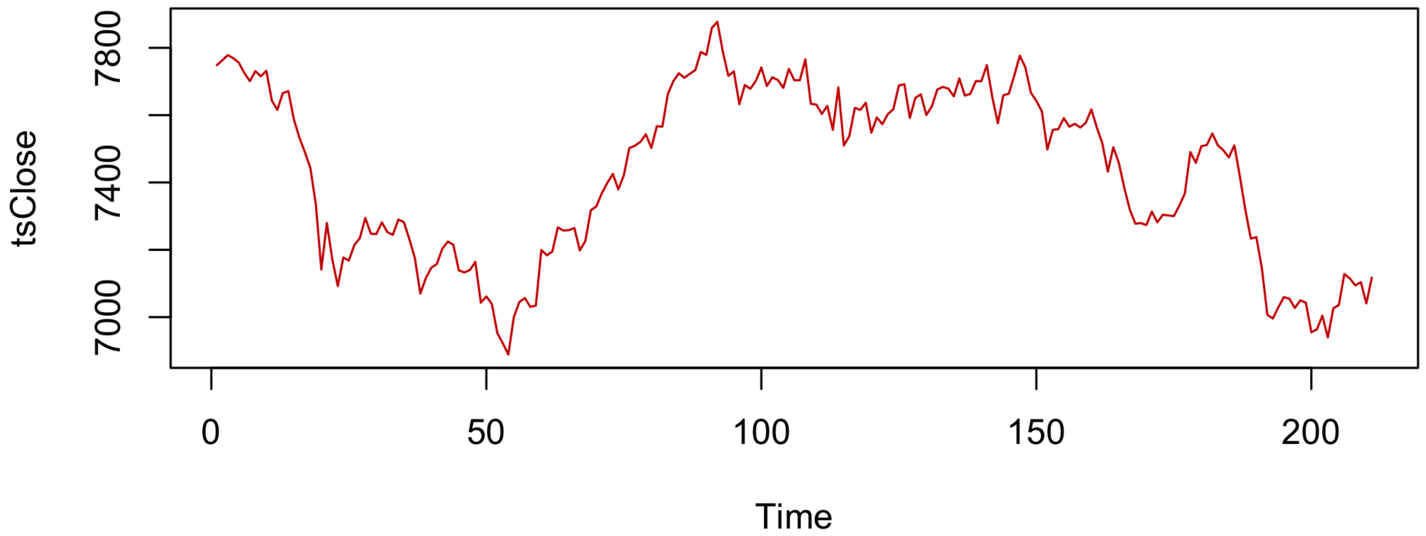
```
##           Date    Open    High     Low   Close
## 211 1/10/2018 7731.02 7756.11 7716.21 7748.51
## 210 1/11/2018 7748.51 7768.96 7734.64 7762.94
## 209 1/12/2018 7762.94 7792.56 7752.63 7778.64
## 208 1/15/2018 7778.64 7783.61 7763.43 7769.14
## 207 1/16/2018 7769.14 7791.83 7740.55 7755.93
## 206 1/17/2018 7755.93 7755.93 7711.11 7725.43
```

IDS 702

Motivating example: FTSE 100

Plot the closing prices to see what a simple time series data looks like.

```
tsClose <- ts(ftse100$Close); ts.plot(tsClose,col="red3")
```



- It is reasonable to expect closing prices for a particular day to be correlated with closing prices for previous days.
- How many of the previous days? We will have to investigate!

IDS 702

Motivating example: Sunspots and melanoma

- We will revisit that data but let's look at different example, where we also have a predictor.
- Incidence of melanoma (skin cancer) may be related to solar radiation.
- Annual data from Connecticut tumor registry on age adjusted melanoma incidence rates (per 100000 people).
- Treat these rates as without error.
- We also have annual data on relative sunspot (dark spots on the sun caused by intense magnetic activity) activity.
- Data go from 1936 to 1972.

Motivating example: Sunspots and melanoma

```r
cancersun <- read.csv("data/melanoma.csv", header = T)
names(cancersun) = c("year", "melanoma", "sunspot")
str(cancersun)
```

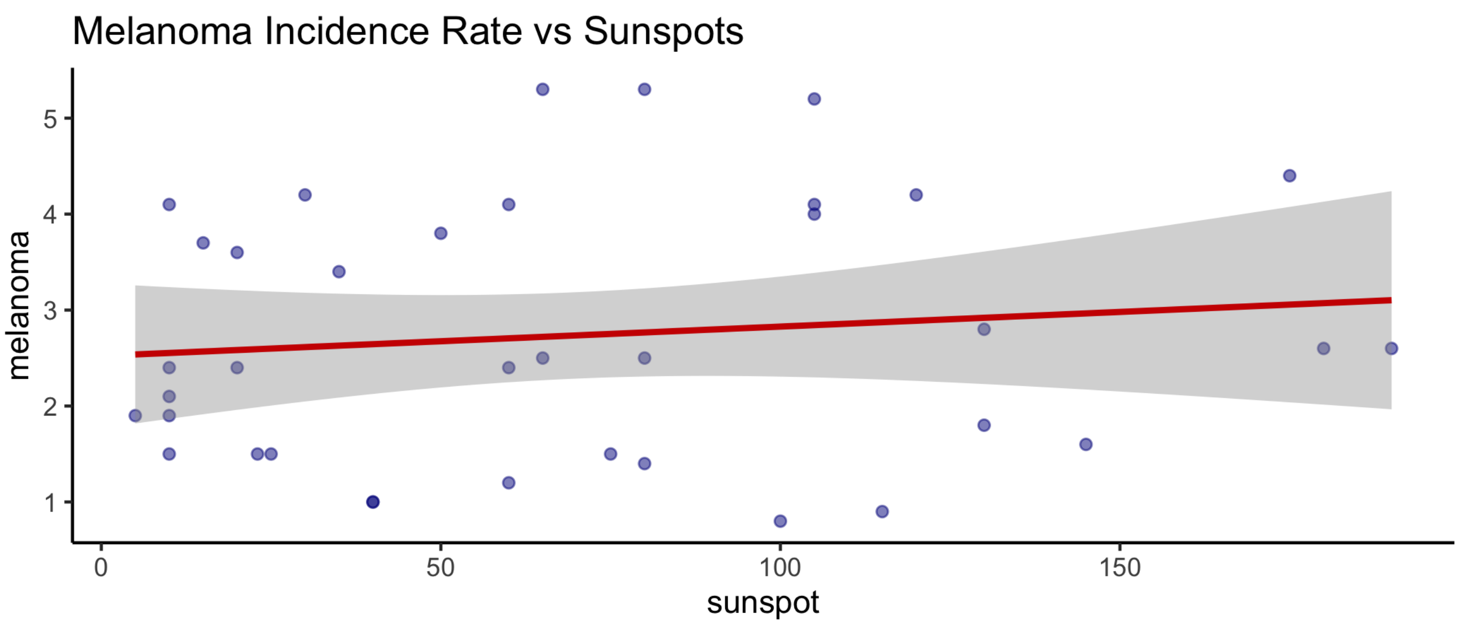
```
## 'data.frame':	37 obs. of  3 variables:
##  $ year    : int  1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 ...
##  $ melanoma: num  1 0.9 0.8 1.4 1.2 1 1.5 1.9 1.5 1.5 ...
##  $ sunspot : num  40 115 100 80 60 40 23 10 10 25 ...
```

```r
head(cancersun)
```

```
##   year melanoma sunspot
## 1 1936      1.0      40
## 2 1937      0.9     115
## 3 1938      0.8     100
## 4 1939      1.4      80
## 5 1940      1.2      60
## 6 1941      1.0      40
```

Motivating example: Sunspots and melanoma

```
ggplot(cancersun, aes(x=sunspot, y=melanoma)) +
  geom_point(alpha = .5,colour="blue4") +
  geom_smooth(method="lm",col="red3") +
  labs(title="Melanoma Incidence Rate vs Sunspots") +
  theme_classic()
```

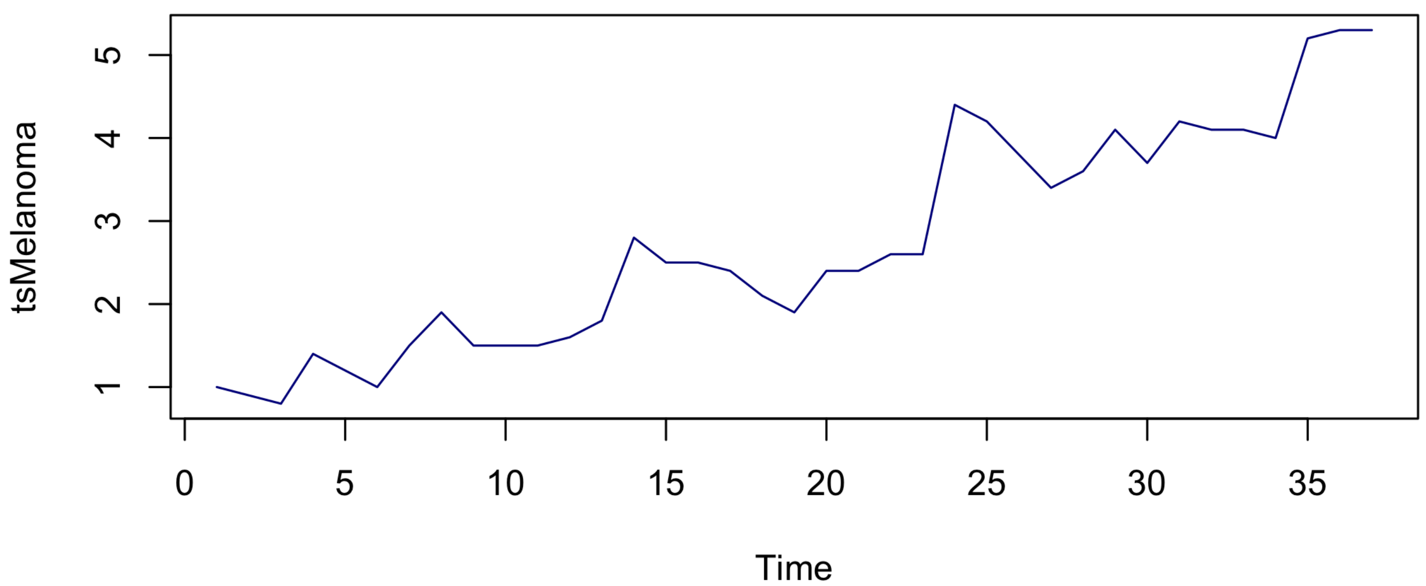


Weak positive (maybe!) relationship between them.

Motivating example: Sunspots and melanoma

Let's look at melanoma incidence rate in time

```
tsMelanoma <- ts(cancersun$melanoma); ts.plot(tsMelanoma,col="blue4")
```



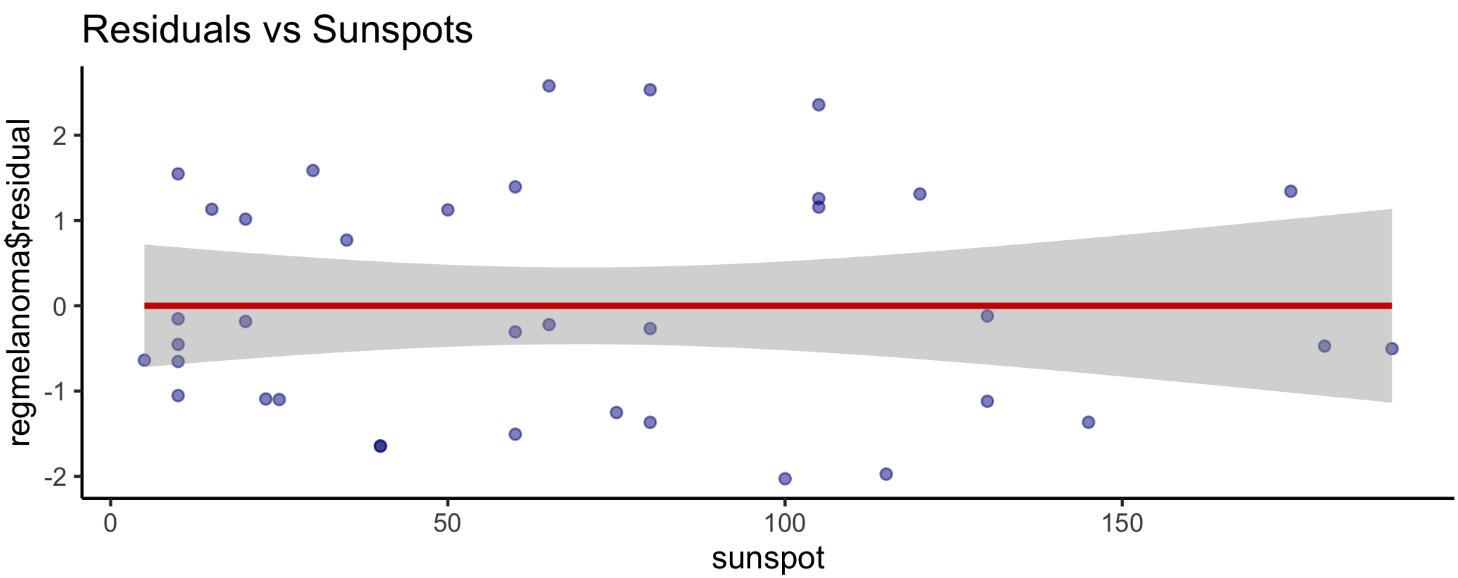
Trend in time, some of which we might be able to explain using `sunspots`.

IDS 702

Motivating example: Sunspots and melanoma

Let's fit a linear model to the relationship between the two variables.

```
regmelanoma = lm(melanoma ~ sunspot, data = cancersun)
ggplot(cancersun, aes(x=sunspot, y=regmelanoma$residual)) +
  geom_point(alpha = .5,colour="blue4") +
  geom_smooth(method="lm",col="red3") + labs(title="Residuals vs Sunspots") +
  theme_classic()
```

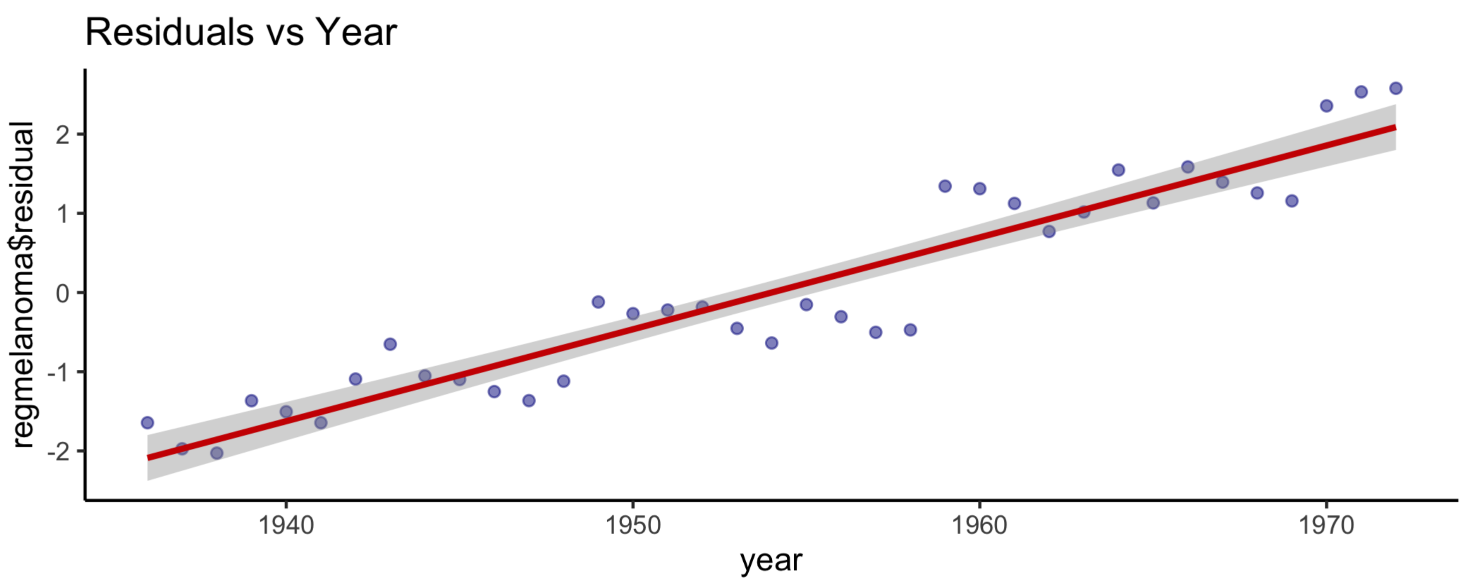


Residuals look fine here.

Motivating example: Sunspots and melanoma

Let's plot the residuals versus year.

```
ggplot(cancersun, aes(x=year, y=regmelanoma$residual)) +
  geom_point(alpha = .5,colour="blue4") +
  geom_smooth(method="lm",col="red3") + labs(title="Residuals vs Year") +
  theme_classic()
```



Huge trend! What to do???

What's next?

Move on to the readings for the next module!

IDS 702