

# Environmental Informatics

Ryan G

February 15, 2020

## Contents

### 1 **DONE** Link to PDF / T<sub>E</sub>X / DocX fileATTACH

This is incomplete because of the conversion, Refer to:

- Original PDF Version
  - DocX File not Yet Attached

### 2 **TODO** Get Working into this document

Either get the working out of the scripts and transcribe into org-mode.

OR

Clean them up and make them into R Scripts/RMD documents in vim and then add them to:

- MD Notes
- Link them to this Org file
- put the ***R**-Markdown* / scripts in the DataSci Project
  - I want to have a git repo with clean exemplar code in it and I want it to be a project, when I start a unit or a semester I want to add the exemplar code as a submodule so I have something to work off, questions I need to consider are:
    - \* How will I tag things?
      - Probably **#tags**
    - \* How will I search tags

- Can I use my old **#tag** script?
- \* How will i integrate this with theory notes from **Org** and MD?
- \* How will I search Documents
  - Probably **FZF** and **RG**
- \* When do i want to use RMD and when do I want to use MD?
- This will be clearer once I:
  1. remove duplicate notable notes
  2. Import TSA Scripts and RMD files

### 3 Original Material

### ATTACH

Chapman TextBook

1. Correlation a. Script
2. Confidence Intervals
  - (a) Notes
    - i. Source
  - (b) Lecture Material
3. Tolerance Intervals 1.Lecture Material
  - (a) Topic Notes
    - i. Word Doc Source
  - (b) Scripts
4. Linear Regression
  - (a) Scripts
  - (b) Working Exercises (PDF)
    - i.  $\text{\LaTeX}$  Source
  - (c) Notes
    - i. Source (DocX)
5. Multiple Linear Regression
  - (a) Notes
    - i. Applied Statistics Notes

- A. Source (MS Word)
  - ii. Relevant Notes
    - A. Source (Ms Word)
  - (b) Lecture material
  - (c) Worked Exercise Solutions
    - i. Scripts
- 6. Time Series Analysis
  - (a) Notes
    - i. Source
  - (b) Old Notes Applied Stats
    - i. A
    - ii. B
  - (c) Exercises Worked Solutions
    - i. Scripts
  - (d) Lecture Material
- 7. Trend Estimation
  - (a) Notes
    - i. Source
  - (b) Exercises Worked Solutions
    - i. Scripts
    - ii. Worked Solutions
      - A. 7.1.1
      - B. 7.1.2
      - C. 7.1.3
      - D. 7.1.4
      - E. 7.1.5
      - F. 7.1.6
      - G. 7.2
  - (c) Lecture Material
- 8. AR, MA, ARMA and ARIMA
  - (a) Notes 1.Source

- (b) Exercises Worked Solutions

- i. 8.1
- ii. 8.2
- iii. 8.3
- iv. Scripts

- (c) Lecture Material

I haven't attached/linked anything beneath here for want of times and concern over file size.

1. Model Selection

- (a) Notes

- i. Source

- (b) Exercises Worked Solutions

- i. Scripts

- (c) Lecture Material

2. GeoSpatial

- (a) Notes

- i. Source

- (b) Exercises Worked Solutions

- i. Scripts

- (c) Lecture Material

3. Variograms

- (a) Notes

- i. Source

- (b) Exercises Worked Solutions

- i. Scripts

- (c) Lecture Material

4. GeoSpatial

- (a) Notes

- i. Source

- (b) Exercises Worked Solutions
    - i. Scripts
  - (c) Lecture Material
- 5. Spatio-Temporal Models
  - (a) Notes
    - i. Source
  - (b) Exercises Worked Solutions
    - i. Scripts
  - (c) Lecture Material

## 4 (1) - Exercises

Week 1 | Material Due: 17 July 2017

### 4.1 Regulations relating to Environmental Hypotheses Testing

Things that require measurement's, ostensibly by law, would include fish populations, ocean acidity, CO<sub>2</sub> levels, temperature, rainfall etc.

Finding which legislative instrument provides for this is difficult, a cursory glance through *Westlaw*, *LexisNexis*, *Google* and *Austlii* does not provide anything obvious.

### 4.2 Summary of Temperature Data

A table of data with rows as observations and columns as variables is a data frame.

#### 4.2.1 First few Lines

#### 4.2.2 Structure of the Data Set

#### 4.2.3 Summary of the Data Frame

#### 4.2.4 Correlation of Minimum and Maximum Temperature

The value provided by the `cor` function is the default *Pearson* method<sup>1</sup> which is a *linear correlation coefficient* such that:<sup>2</sup>

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \times \sqrt{n(\sum y^2) - (\sum y)^2}}$$

The value of  $r$  measures the strength and direction of a linear relationship by comparing the variation, whereby a value of 1 would occur if all the data were to lie exactly on a straight line, a strong correlation is usually  $r > \pm 0.8$ .

The Coefficient of determination is  $r^2$ :

$$r^2 = 0.7498076^2 = 0.562211$$

The Coefficient of determination is the percentage of variation that is explained by the linear function, so in this case 56% of the variation between maximum and minimum temperatures can be explained by a linear function between those two variables, the linear function can be found by using an RSS method, explained by `.Simple\ Linear\ Regression.pdf`.

This is the same method used with Excel:

[CHART]

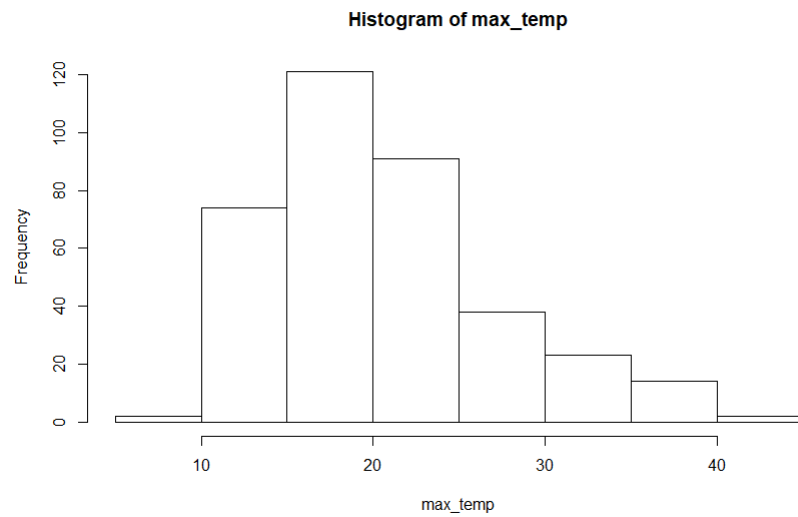
---

<sup>1</sup>Rdocumentation.org. (2017).`/cor` function | R Documentation/. [online] Available at: <https://www.rdocumentation.org/packages/stats/versions/3.4.1/topics/cor> [Accessed 21 Jul. 2017].

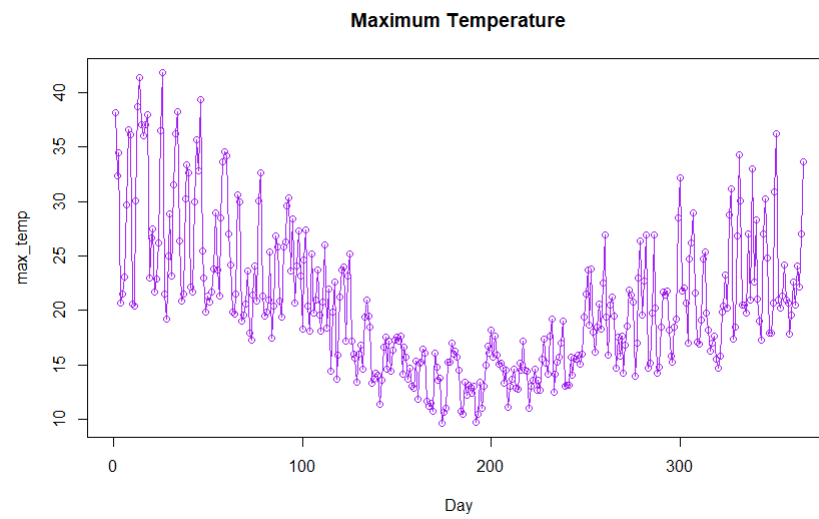
<sup>2</sup>Roberts, D. (2017).`/Statistics 2 - Correlation Coefficient and Coefficient of Determination/`. [online] Mathbits.com. Available at: <https://mathbits.com/MathBits/TISection/Statistics2/correlation.htm> [Accessed 21 Jul. 2017].

## 4.3 1.5 Re-Produce Figures<sup>3</sup>

### 4.3.1 Histogram



### 4.3.2 Line Plot



---

<sup>3</sup>Use this cite as a man page:  
[https://www.tutorialspoint.com/r/r\\_line\\_graphs.htm](https://www.tutorialspoint.com/r/r_line_graphs.htm)