# HW2_arflowers

Anna Flowers

9/8/2021

## Homework 2

### Problem 2

#### Part A

I am excited for this class because although I have been working with R for about 3 years now, I have never taken a class specifically designed to teach it. The knowledge I have of R has come from classes that used it to accompany the material (so any teaching revolved around that class), or from internet sources that I used to teach myself. I am especially excited to learn more about LaTeX, because although I have used it in the past I have relied on the support of others to create a coherent document. I also want to experience more of the combination of R and LaTeX, because in the past I have really only used LaTeX through Overleaf in documents entirely in LaTeX. I also want to learn more about Github, because although I have used it in the past I have exclusively relied on repositories used by other people and am not used to creating my own material yet.

1. Learn more efficient ways to accomplish tasks I may have been over complicating.

2. Become more confident in my skills using LaTeX.

3. Become more comfortable using Github with repositories that I have created myself.

#### Part B

Gamma Density Function:

$$f\left(x|\alpha,\beta\right) = \frac{1}{\Gamma\left(\alpha\right)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta}; 0 \leq x < \infty; \alpha,\beta > 0 \tag{1}$$

Chi squared Density Function:

$$f\left(x|p\right) = \frac{1}{\Gamma\left(p/2\right)2^{p/2}}x^{(p/2)-1}e^{-x/2}; 0 \leq x < \infty; p = 1,2,... \tag{2}$$

Lognormal Density Function:

$$f\left(x|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log x-\mu)^2/\left(2\sigma^2\right)}}{x}; 0 \leq x < \infty; -\infty < \mu < \infty \tag{3}$$

## Problem 3

1. Keep track of how all results were produced

   - Challenges: These results are often produced through a lot of trial and error (especially when coding is involved), so it can be difficult to separate the steps that were important and those that were unnecessary.

2. Avoid manually manipulating data

   - Challenges: Some machines may not be able to read specific data types, but opening the file with a different type may change the data without your knowledge.

3. Archive (or keep track of) the exact versions of all programs used

   - Challenges: Even if you do archive the program, some updates will automatically delete older versions on the computer, possibly rendering past research useless.

4. Version control all scripts

   - Challenges: It can be difficult to know when to store versions of code and when not to. Too many versions of code can still make the correct version difficult to find, and too few means you are less likely to have the exact code that you want.

5. Record all intermediate results and standardize if possible

   - Challenges:

6. Note seeds used for analyses that include randomness

   - Challenges:

7. Store raw data used to make plots

   - Challenges:

8. Keep and inspect all layers of detail of the data

   - Challenges:

9. Connect statements to the results that inspired them

   - Challenges:

10. Provide public access to all data used, programs written, and results discovered

    - Challenges:

## Problem 4

```
#install.packages('data.table')
library(data.table)
covid_raw <- fread("https://opendata.ecdc.europa.eu/covid19/casedistribution/csv")
us <- covid_raw[covid_raw$countriesAndTerritories == 'United_States_of_America',]
us_filtered <- us[us$month %in% c(6:7),]
us_filtered$index <- rev(1:dim(us_filtered)[1])
fit<-lm(`Cumulative_number_for_14_days_of_COVID-19_cases_per_100000`~index, data=us_filtered)
```

### Part A

```
library(knitr)
kable(summary(us_filtered))
```

| | dateRep | day | month | year | cases | deaths | countriesAndTerritories | geoId | countryterritoryCode | popData2019 | continentExp | Cumulative_number_for_14_days_of_COVID-19_cases_per_100000 | index |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Length:61 | Min. : 1.00 | Min. :6.000 | Min. :2020 | Min. :18665 | Min. : 242.0 | Length:61 | Length:61 | Length:61 | Min. :329064917 | Length:61 | Min. : 89.76 | Min. : 1 |
| | Class :character | 1st Qu.: 8.00 | 1st Qu.:6.000 | 1st Qu.:2020 | 1st Qu.:25640 | 1st Qu.: 500.0 | Class :character | Class :character | Class :character | 1st Qu.:329064917 | Class :character | 1st Qu.: 92.43 | 1st Qu.:16 |
| | Mode :character | Median :16.00 | Median :7.000 | Median :2020 | Median :45221 | Median : 767.0 | Mode :character | Mode :character | Mode :character | Median :329064917 | Mode :character | Median :150.94 | Median :31 |
| | NA | Mean :15.75 | Mean :6.508 | Mean :2020 | Mean :44666 | Mean : 791.6 | NA | NA | NA | Mean :329064917 | NA | Mean :170.16 | Mean :31 |
| | NA | 3rd Qu.:23.00 | 3rd Qu.:7.000 | 3rd Qu.:2020 | 3rd Qu.:61796 | 3rd Qu.: 982.0 | NA | NA | NA | 3rd Qu.:329064917 | NA | 3rd Qu.:247.01 | 3rd Qu.:46 |
| | NA | Max. :31.00 | Max. :7.000 | Max. :2020 | Max. :78427 | Max. :2437.0 | NA | NA | NA | Max. :329064917 | NA | Max. :282.72 | Max. :61 |

This data is limited to 61 time points from June 2020 to July 2020. There are no missing points, since there are 30 days in June and 31 in July, so that gives a total of 61 days to survey.

```
library(stargazer)
```

```
##
## Please cite as:
##   Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##   R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```
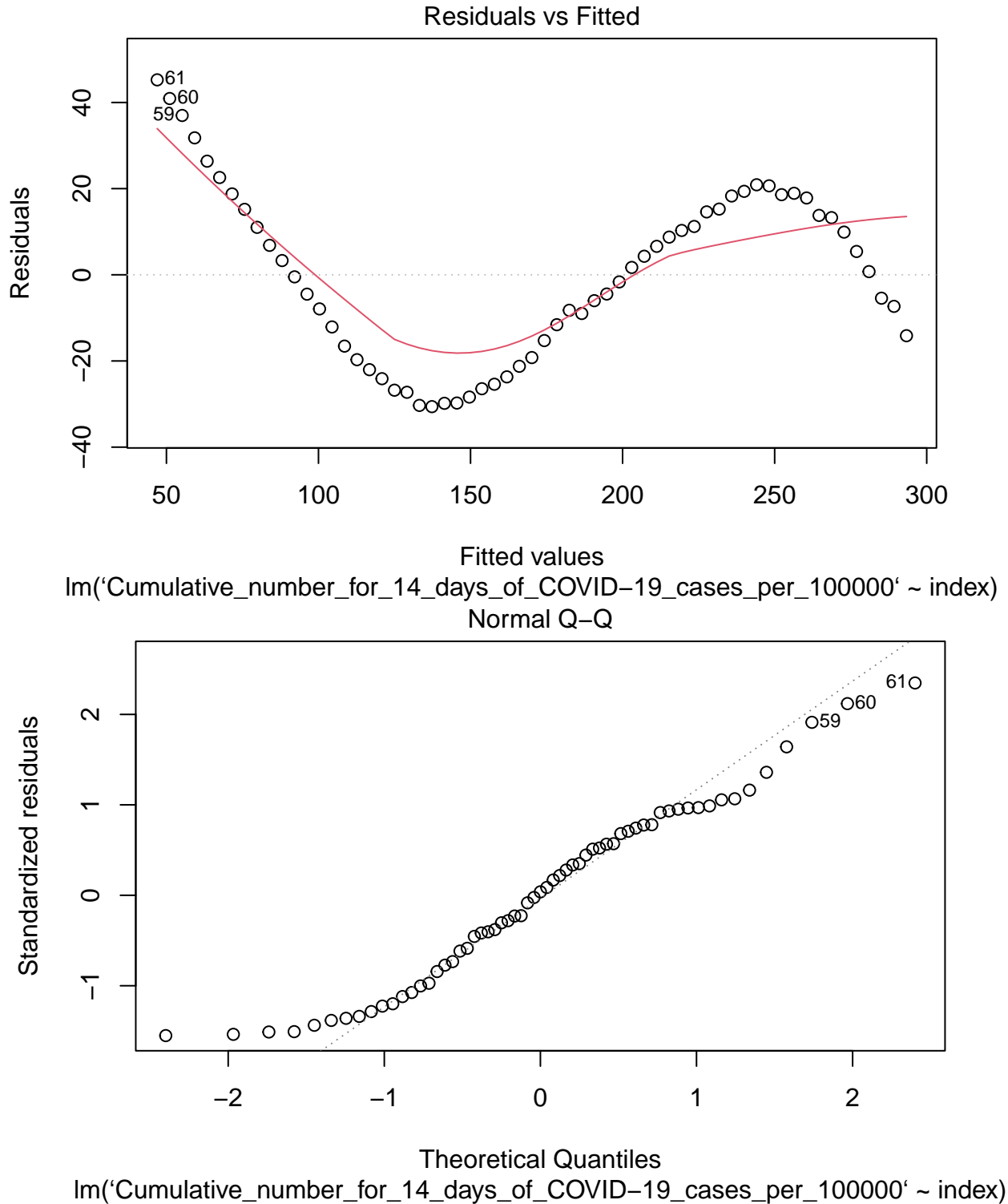
```
#stargazer(fit)
```

Table 2:

| | *Dependent variable:* |
|---|---|
| | 'Cumulative_number_for_14_days_of_COVID-19_cases_per_100000' |
| index | 4.107*** |
| | (0.145) |
| | |
| Constant | 42.853*** |
| | (5.165) |
| | |
| Observations | 61 |
| $R^2$ | 0.932 |
| Adjusted $R^2$ | 0.930 |
| Residual Std. Error | 19.922 (df = 59) |
| F Statistic | 803.464*** (df = 1; 59) |

| Note: | $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |
|---|---|

**Part B**

```r
#install.packages("broom")
fit.diags <- broom::augment(fit)
plot(fit,c(1:3,5))
```

### Residuals vs Fitted



Fitted values
lm('Cumulative_number_for_14_days_of_COVID−19_cases_per_100000' ~ index)

### Normal Q–Q



Theoretical Quantiles
lm('Cumulative_number_for_14_days_of_COVID−19_cases_per_100000' ~ index)

**Scale–Location**

lm(‘Cumulative_number_for_14_days_of_COVID−19_cases_per_100000‘ ~ index)
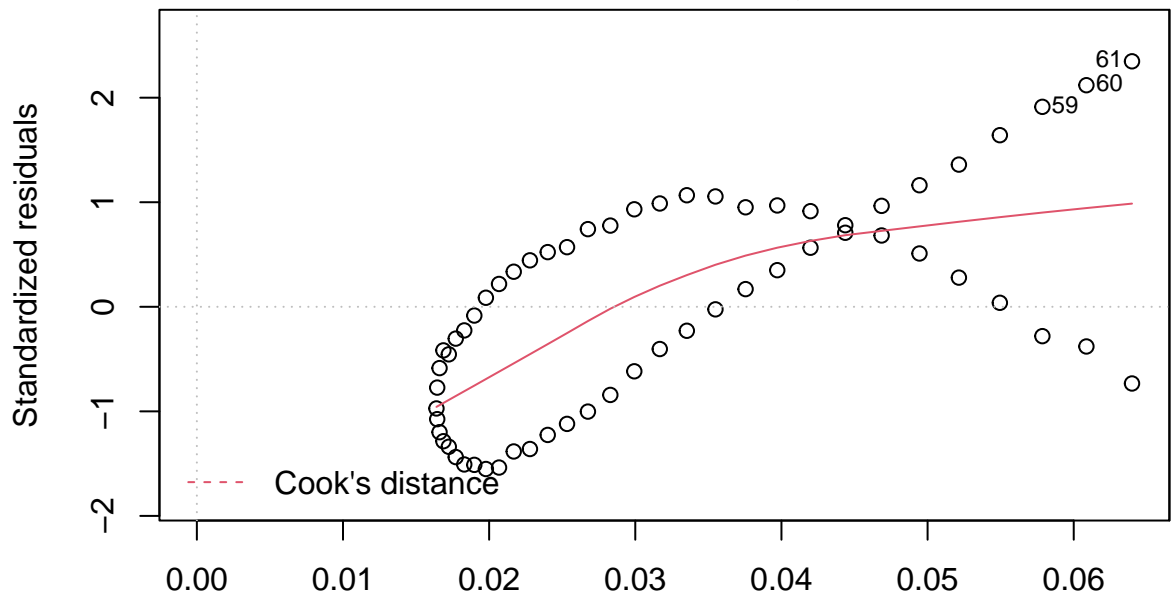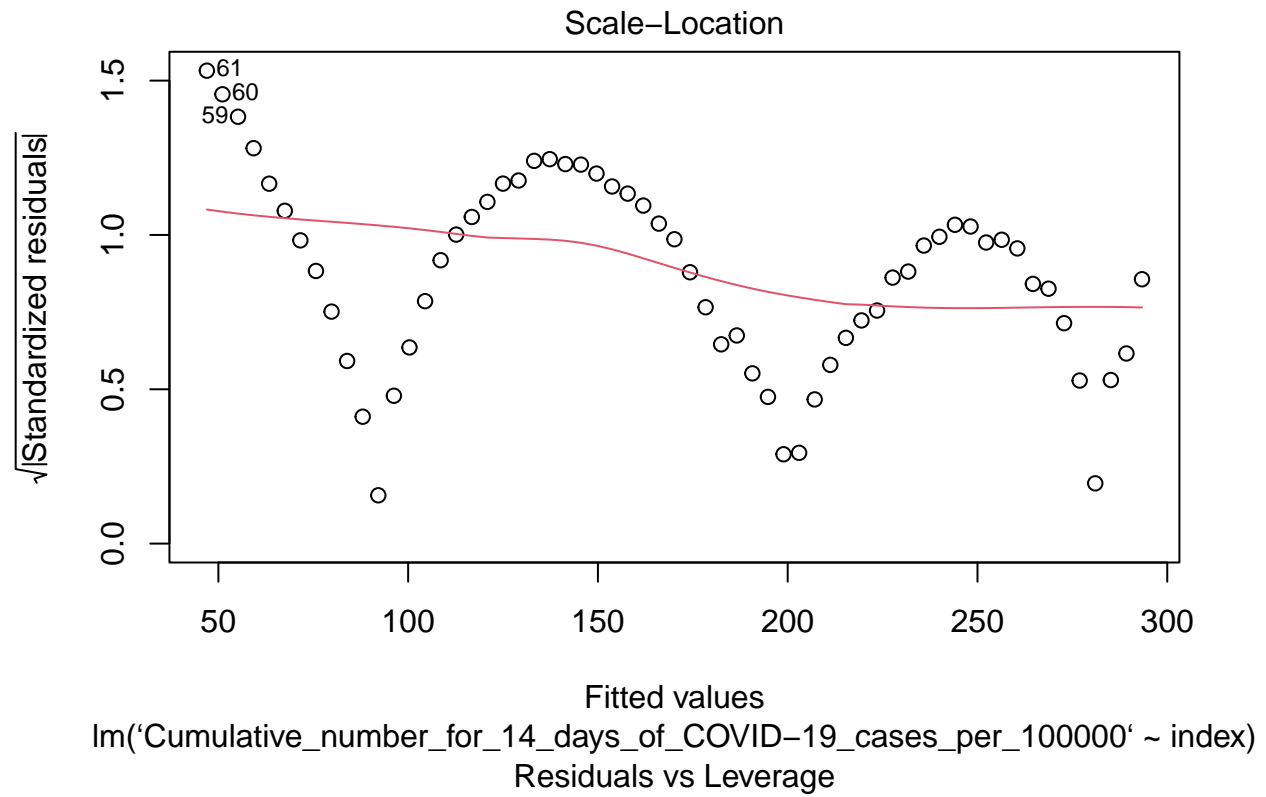


**Residuals vs Leverage**
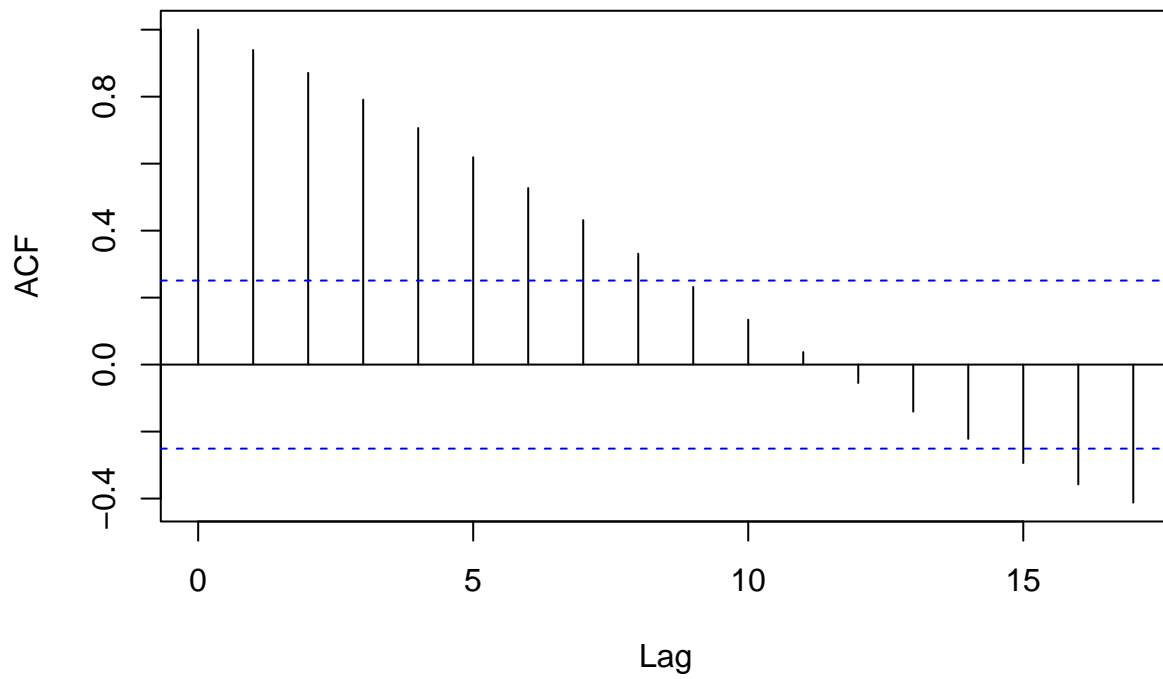
lm(‘Cumulative_number_for_14_days_of_COVID−19_cases_per_100000‘ ~ index)

**Part C**

```
acf(fit$residuals)
```

## Series  fit$residuals



## Problem 5

```
par(mfrow=c(2,2))
plot(fit, c(1:3,5))
```