HW2_arflowers

Anna Flowers

9/8/2021

Homework 2

Problem 2

Part A

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

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.:25040	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
```

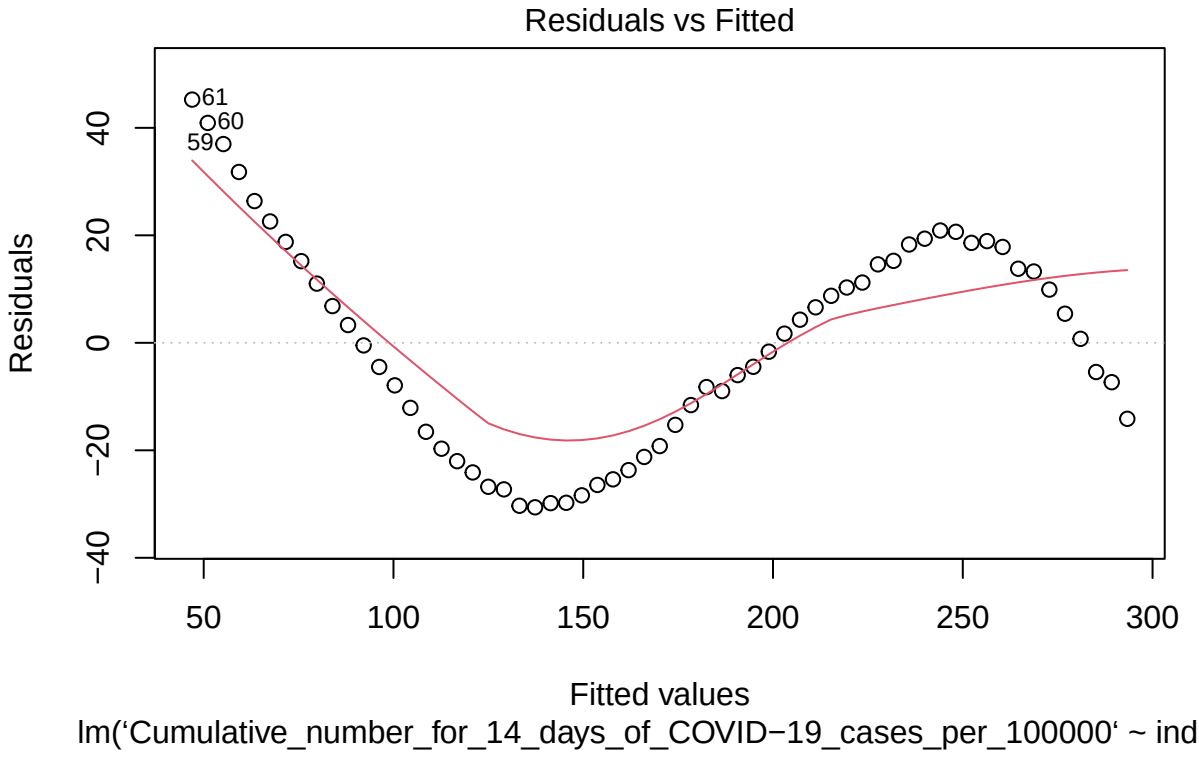
```
#stargazer(fit)
```

Table 2:

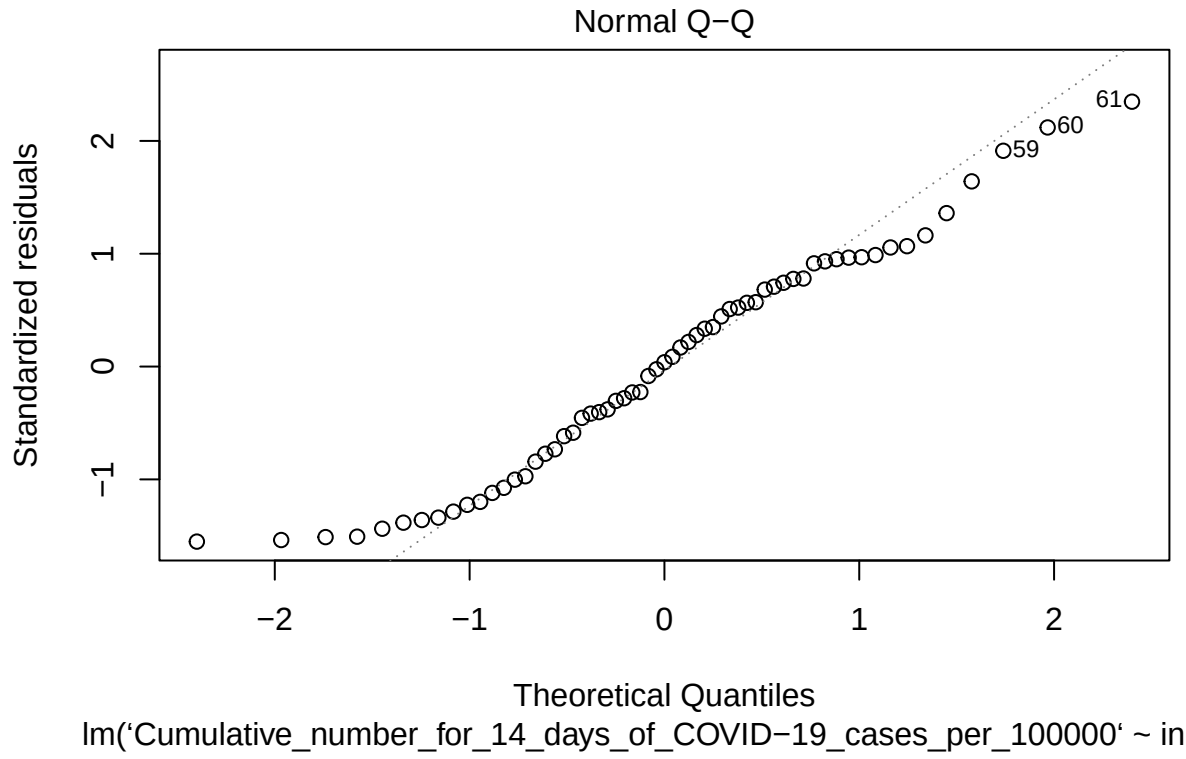
	Dependent variable:
	‘Cumulative_number_for_14_days_of_COVID-19_cases_per_100000‘
index	4.107***
	(0.145)
Constant	42.853***
	(5.165)
Observations	61
R^2	0.932
Adjusted R^2	0.930
Residual Std. Error	19.922 (df = 59)
F Statistic	803.464*** (df = 1; 59)
Note:	*p<0.1; **p<0.05; ***p<0.01

Part B

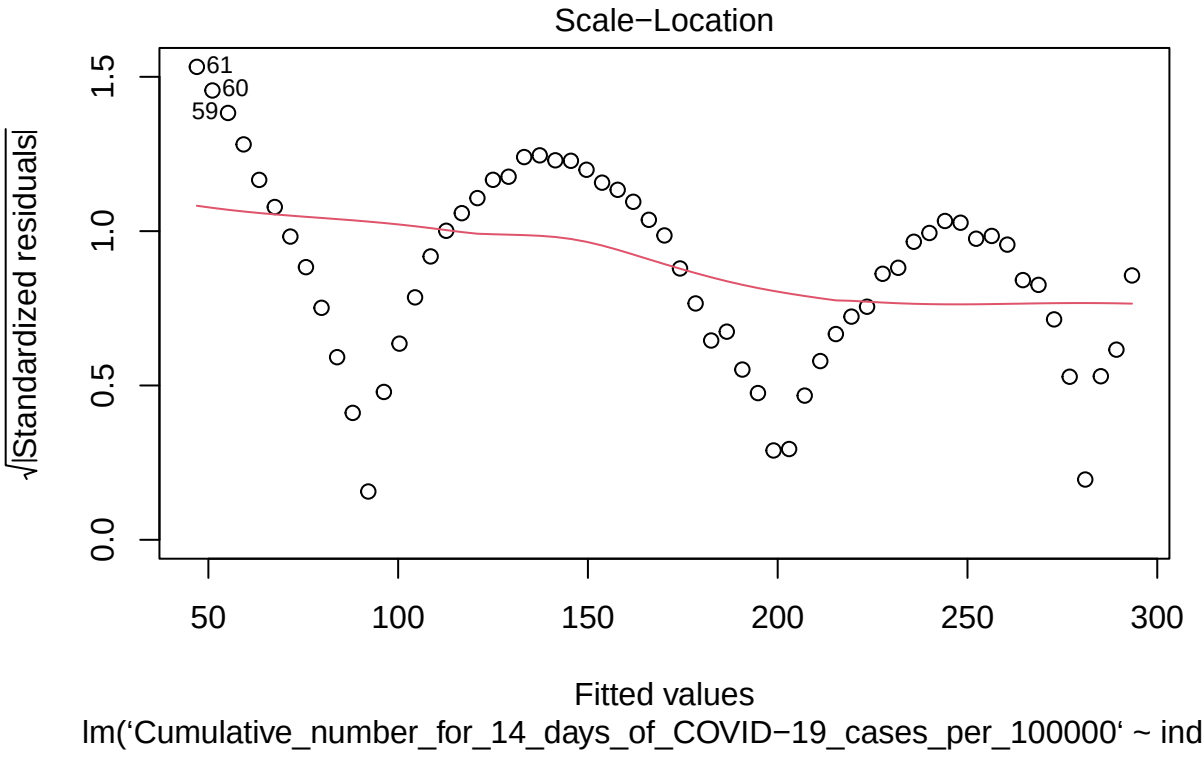
```
#install.packages("broom")
fit.diags <- broom::augment(fit)
plot(fit,1)
```



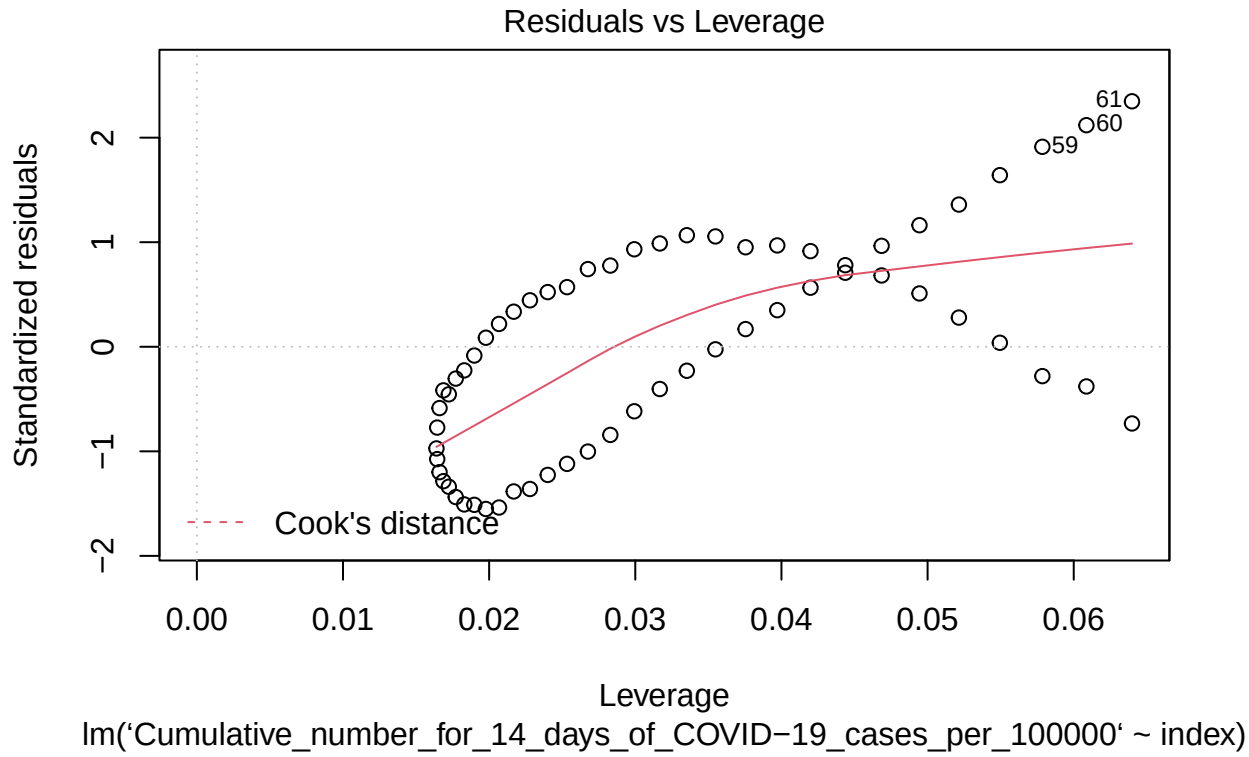

```
plot(fit,2)
```




```
plot(fit,3)
```

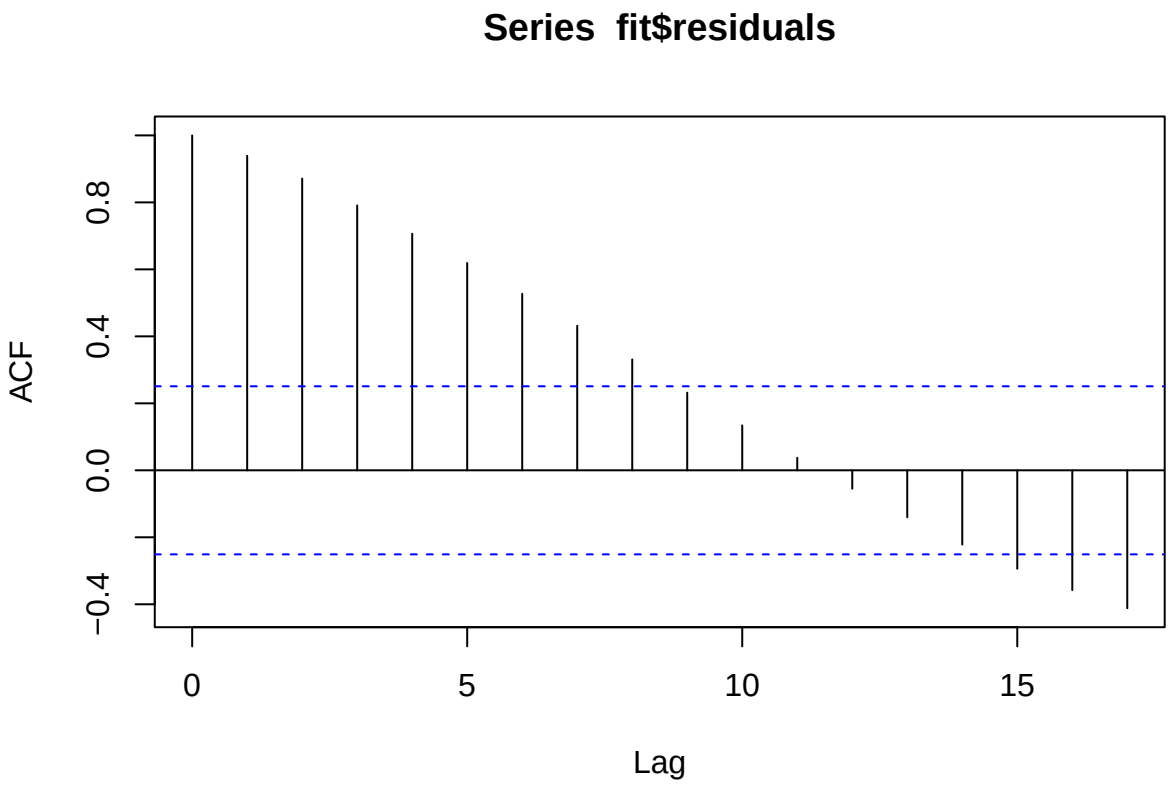



```
plot(fit,5)
```



Part C

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



Problem 5

Problem 6