코로나 확진자 수의 선형 및 비선형 회귀 모형 적합

Son Jaehyeon

2021 4 29

1. Introduction

본 보고서에서는 코로나 발발 이후 시점부터의 누적 확진자 수를 적절한 회귀 모형으로 적합한다. 이후, 백신 접종이 일일 코로나 확진자 수 감소에 기여하고 있는지 extra sum of square 방식으로 검정하고자 한다. 국가들 중 러시아의 데이터를 사용하여 분석을 진행하였다.

소프트웨어는 R 4.0.2 version을 사용하였다. 사용할 라이브러리는 다음과 같다.

```
library(dplyr)
library(ggplot2)
library(nls2)
library(segmented)
```

2. EDA & Data Preprocessing

2.1 Cumulative Cases Fitting

```
df <- read.csv("global_confirmed_cases_210420.csv")
str(df)
```

```
## 'data.frame':    75148 obs. of  6 variables:
##  $ CountryName: chr  "Aruba" "Aruba" "Aruba" "Aruba" ...
##  $ CountryCode: chr  "ABW" "ABW" "ABW" "ABW" ...
##  $ Date       : chr  "2020.3.13" "2020.3.14" "2020.3.15" "2020.3.16" ...
##  $ Cases      : int  2 2 2 2 3 4 4 5 5 9 ...
##  $ Difference : int  2 0 0 0 1 1 0 1 0 4 ...
##  $ Days       : int  1 2 3 4 5 6 7 8 9 10 ...
```

데이터를 로드하였다. 이 중 러시아 데이터만을 추출하여 적절히 변수형을 변환하였다.

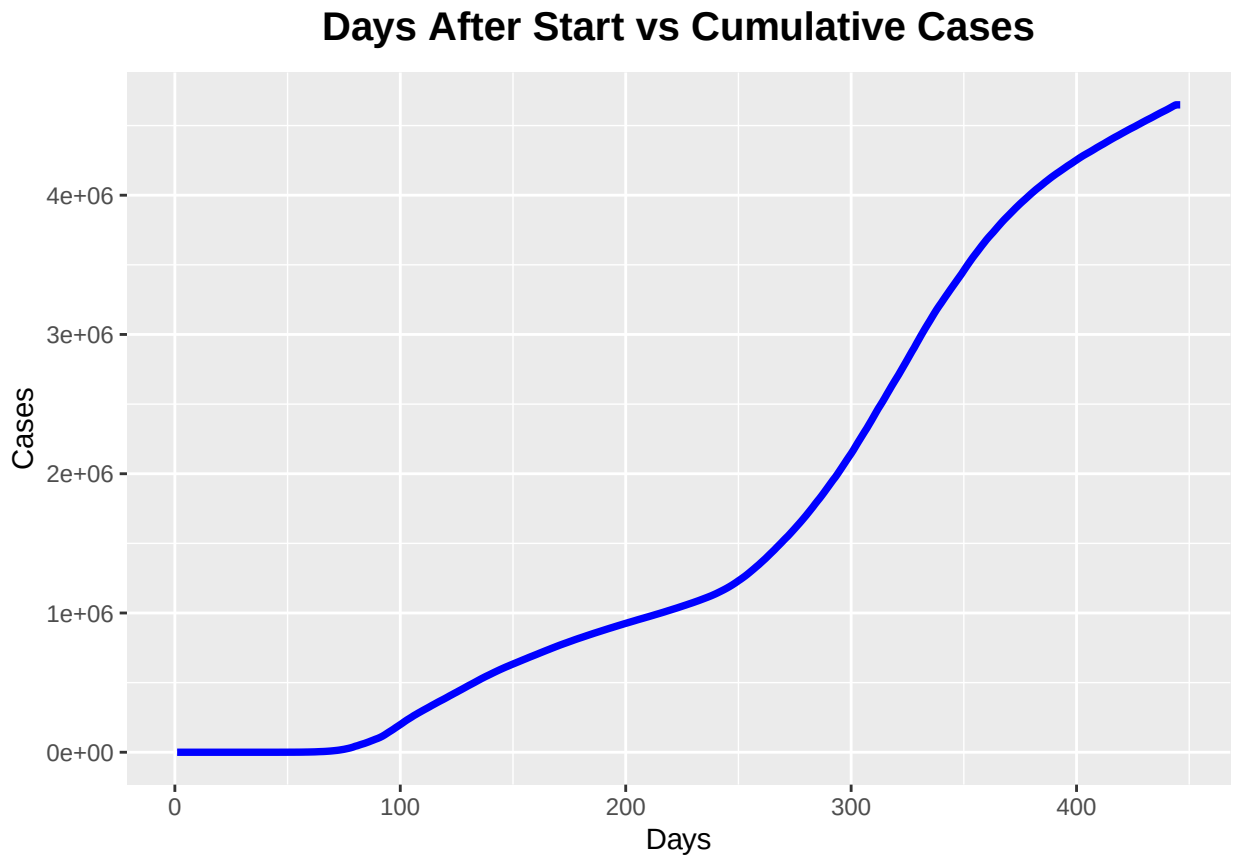
```
df_RUS <- filter(df, CountryCode == "RUS")
df_RUS$Date <- df_RUS$Date %>% as.Date("%Y.%m.%d")
```

```
head(df_RUS)
```

```
##   CountryName CountryCode       Date Cases Difference Days
## 1      Russia         RUS 2020-01-31     2          2    1
## 2      Russia         RUS 2020-02-01     2          0    2
## 3      Russia         RUS 2020-02-02     2          0    3
## 4      Russia         RUS 2020-02-03     2          0    4
## 5      Russia         RUS 2020-02-04     2          0    5
## 6      Russia         RUS 2020-02-05     2          0    6
```

먼저 위의 Cases 변수를 response로, Days 변수를 regressor로 적합하기 위해 두 변수를 간단히 시각화 하였다.

```
ggplot(data = df_RUS, mapping = aes(Days, Cases)) +
  geom_line(size = 1.3, colour = 'blue') +
  ggtitle("Days After Start vs Cumulative Cases") +
  theme(plot.title=element_text(size=15, hjust=0.5, face="bold", colour="black", vjust=2))
```



시각화 결과를 바탕으로, 선형 모형 중 4~6차 정도의 polynomial model을 적합하고, 비선형 모형 중 logistic model과 Gompertz model을 적합하여 서로 비교하는 방식으로 분석하기로 결정하였다.

이후 overfitting을 방지하기 위하여 데이터를 train set과 test set으로 나누었다. 데이터 중 마지막 1개월만을 test set으로 분리하였다.

```
df_RUS_train <- df_RUS %>% filter(Date <= "2021-03-20")
df_RUS_test <- df_RUS %>% filter(Date > "2021-03-20")
```

2.2 Effect of Vaccine on Daily Cases

누적 백신 접종자 수가 추가로 포함된 데이터를 로드하여 마찬가지로 전처리하였다. 또한 최초 백신 접종일부터 지난 날짜를 Days 변수로 지정하였다.

```
vacc <- read.csv("covid_vaccine.csv")
str(vacc)
```

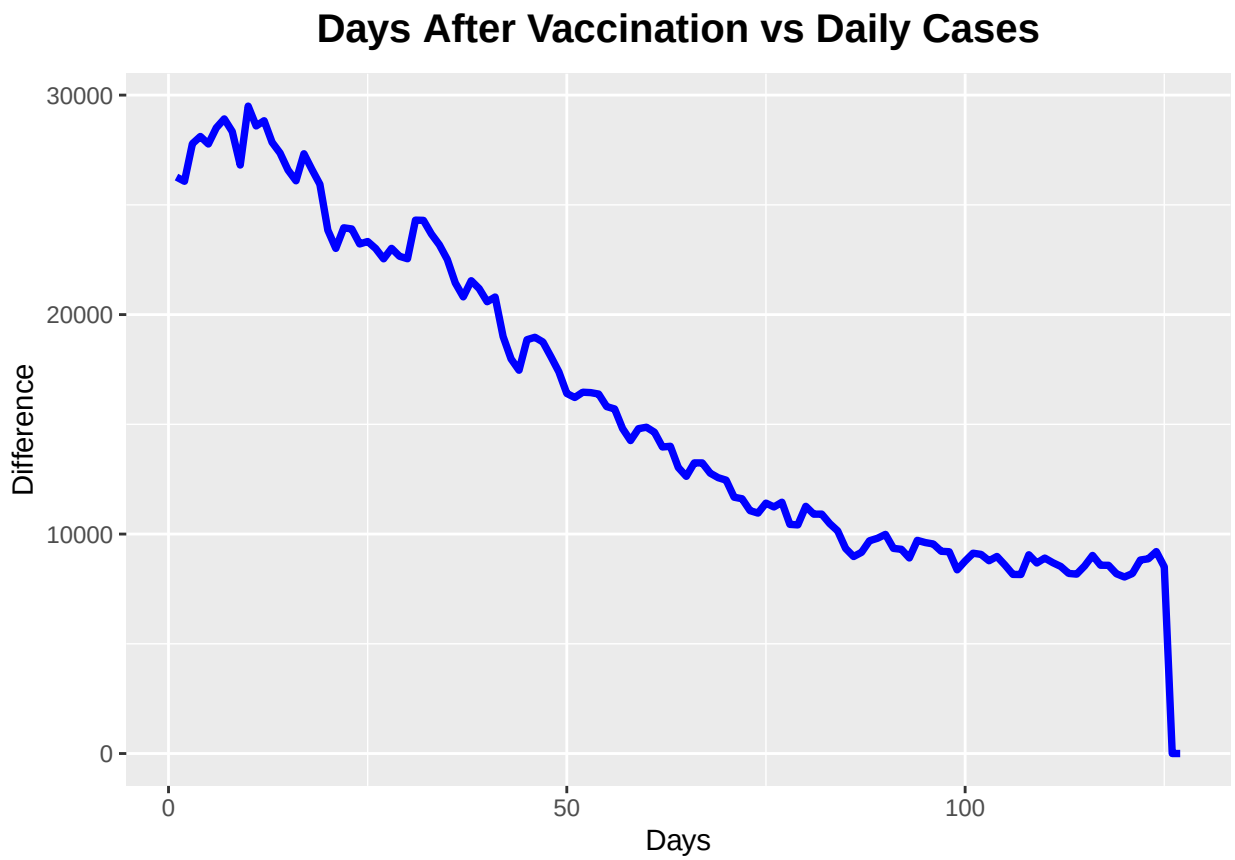
```
## 'data.frame':    3121 obs. of  6 variables:
##  $ CountryName      : chr  "United States" "United States" "United States" "United States" ...
##  $ CountryCode      : chr  "USA" "USA" "USA" "USA" ...
##  $ Date             : chr  "2020.12.20" "2020.12.21" "2020.12.22" "2020.12.23" ...
##  $ Cases            : int  17954675 18153724 18351735 18581353 18775557 18873203 19099491 19255126 19
##  $ Difference       : int  187819 199049 198011 229618 194204 97646 226288 155635 174634 200252 ...
##  $ people_vaccinated: int  556208 614117 614117 1008025 1008025 1008025 1944585 1944585 2127143 2127
```

```
vacc_RUS <- vacc %>% filter(CountryCode == "RUS")
vacc_RUS$Date <- vacc_RUS$Date %>% as.Date("%Y.%m.%d")
vacc_RUS <- vacc_RUS %>% mutate(Days = as.integer(Date - vacc_RUS$Date[1] + 1))
head(vacc_RUS)
```

```
##   CountryName CountryCode       Date   Cases Difference people_vaccinated Days
## 1      Russia         RUS 2020-12-15 2682866      26265             28500    1
## 2      Russia         RUS 2020-12-16 2708940      26074             28500    2
## 3      Russia         RUS 2020-12-17 2736727      27787             28500    3
## 4      Russia         RUS 2020-12-18 2764843      28116             28500    4
## 5      Russia         RUS 2020-12-19 2792615      27772             28500    5
## 6      Russia         RUS 2020-12-20 2821125      28510             28500    6
```

Daily Cases와 Days 변수를 시각화하였다.

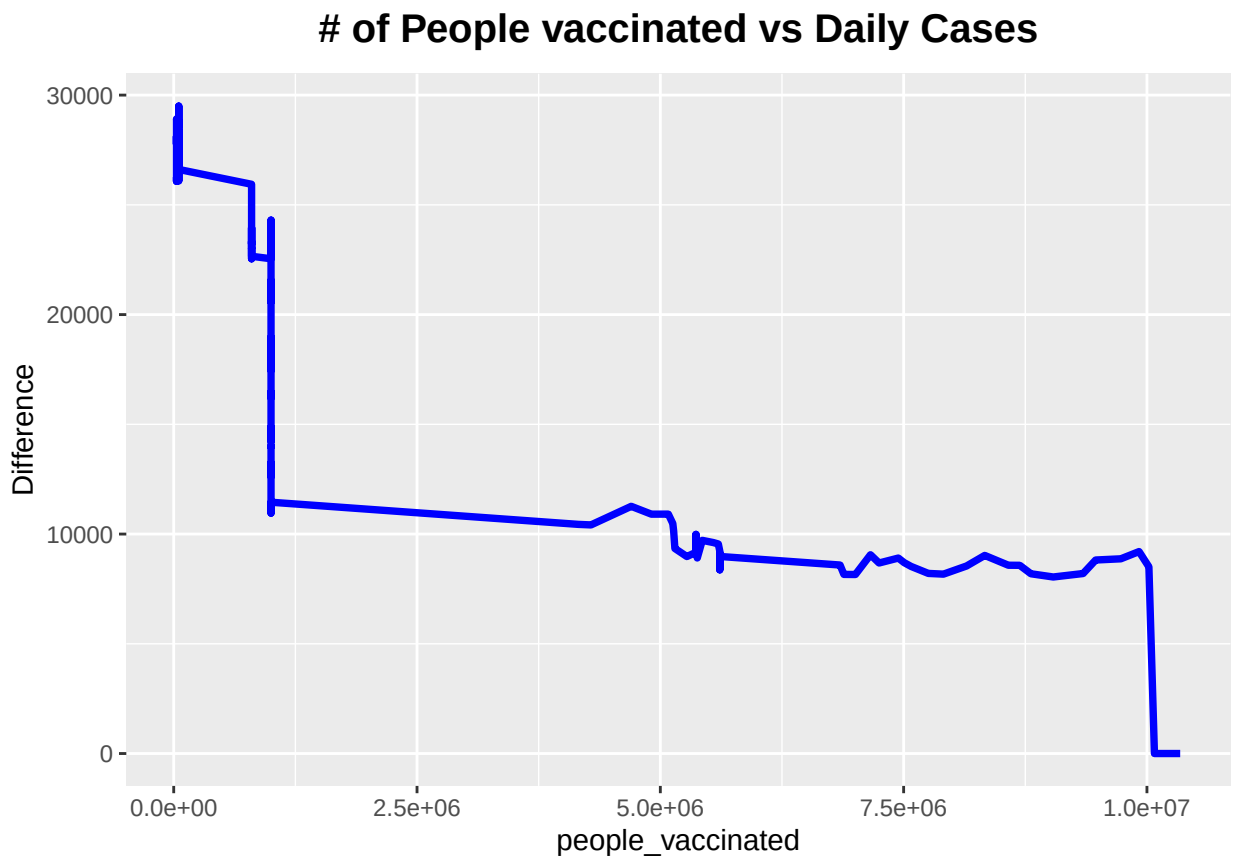
```
ggplot(data = vacc_RUS, mapping = aes(Days, Difference)) +
  geom_line(size = 1.3, colour = 'blue') +
  ggtitle("Days After Vaccination vs Daily Cases") +
  theme(plot.title=element_text(size=15, hjust=0.5, face="bold", colour="black", vjust=2))
```



시간이 지남에 따라 확진자 수가 감소하는 경향을 확인할 수 있었다.

다음으로, 일일 확진자 수와 누적 백신 접종자 수를 시각화하였다.

```
ggplot(data = vacc_RUS, mapping = aes(people_vaccinated, Difference)) +
  geom_line(size = 1.3, colour = 'blue') +
  ggtitle("# of People vaccinated vs Daily Cases") +
  theme(plot.title=element_text(size=15, hjust=0.5, face="bold", colour="black", vjust=2))
```



백신 접종자 수가 늘어남에 따라 일일 확진자 수가 감소하는 경향을 확인할 수 있었다. 그러나, 백신 접종자 수가 충분히 유의하게 일일 확진자 수 감소에 영향을 미쳤는지는 검정이 필요하다고 판단하였다.

train set과 test set을 나누었다. 최근 한달의 데이터만 test set으로 분리해 적합한 모델을 검증할 수 있도록록 하였다.

```
vacc_RUS_train <- vacc_RUS %>% filter(Date <= "2021-03-20")
vacc_RUS_test <- vacc_RUS %>% filter(Date > "2021-03-20")
```

3. Analysis

3.1 Cumulative Cases

3.1.1 Polynomial Model

먼저 누적 확진자 수를 적합하기 위해 polynomial model을 적합하였다. train set으로 적합하여 test set에서 가장 작은 mse를 나타내는 7차 모형을 선택하였다.

```
getMSE <- function(y, yhat){
  MSE <- sum((y-yhat)^2)/length(y)
  return(MSE)
}
```

```
poly.mse <- c()
for(i in 3:8){
  poly <- lm(Cases ~ poly(Days,i), data = df_RUS_train)
  poly.mse <- c(poly.mse, getMSE(df_RUS_test$Cases, predict(poly, df_RUS_test)))
}
print(poly.mse)
```

```
## [1] 6.347075e+11 2.192754e+11 4.529044e+11 1.062444e+12 2.489695e+10
## [6] 3.876874e+11
```

```
linear.fit <- lm(Cases ~ poly(Days,7), data = df_RUS_train)
```

3.1.2 Logistic Model

logistic model을 적합하였다. 적당한 초기값의 grid를 잡아 brute-force 방식으로 모델을 적합한 후, 가우스 뉴턴 방식으로 모델을 최적화하였다.

```
logistic.formula <- Cases ~ a / (1 + exp(b-(c * Days)))
logistic.grid <- expand.grid(a = max(df_RUS_train$Cases),
                             b = seq(1, 100, 1),
                             c = seq(0.01, 1, 0.01))
logistic.fit1 <- nls2(logistic.formula, data = df_RUS_train, start = logistic.grid,
                      algorithm = "brute-force")
logistic.fit2 <- nls2(logistic.formula, data = df_RUS_train, start = coef(logistic.fit1),
                      algorithm = "default")
summary(logistic.fit2)
```

```
##
## Formula: Cases ~ a/(1 + exp(b - (c * Days)))
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## a 6.550e+06  1.610e+05   40.69   <2e-16 ***
## b 4.610e+00  3.735e-02  123.44   <2e-16 ***
## c 1.322e-02  2.293e-04   57.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 131200 on 412 degrees of freedom
##
## Number of iterations to convergence: 7
```

```
## Achieved convergence tolerance: 5.63e-06
```

3.1.3 Gompertz Model

Gompertz model을 적합하였다. 적당한 초기값의 grid를 잡아 brute-force 방식으로 모델을 적합한 후, 가우스 뉴턴 방식으로 모델을 최적화하였다.

```
Gompertz.formula <- Cases ~ a * exp(-b * exp(-c * Days))
Gompertz.grid <- expand.grid(a = max(df_RUS_train$Cases),
                             b = seq(1, 100, 1),
                             c = seq(0, 1, 0.01))
Gompertz.fit1 <- nls2(Gompertz.formula, data = df_RUS_train, start = Gompertz.grid,
                      algorithm = "brute-force")
Gompertz.fit2 <- nls2(Gompertz.formula, data = df_RUS, start = coef(Gompertz.fit1),
                      algorithm = "default")
summary(Gompertz.fit2)
```

```
##
## Formula: Cases ~ a * exp(-b * exp(-c * Days))
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## a 1.021e+07  4.944e+05   20.64   <2e-16 ***
## b 6.810e+00  1.542e-01   44.17   <2e-16 ***
## c 5.066e-03  1.833e-04   27.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168200 on 443 degrees of freedom
##
## Number of iterations to convergence: 10
## Achieved convergence tolerance: 9.866e-06
```

3.1.4 Predict

각 모델의 누적 확진자 수 예측치를 출력하였다. 2021년 3월 20일 까지의 값은 train set에 해당하고, 3월 21일 이후 값은 test set에 해당한다. test dataset의 예측치를 출력하였다.

```
model.linear = linear.fit
model.logistic = logistic.fit2
model.Gompertz = Gompertz.fit2
```

```r
y_hat_linear = predict(model.linear, df_RUS)
y_hat_logistic = predict(model.logistic, df_RUS)
y_hat_Gompertz = predict(model.Gompertz, df_RUS)

predict <- data.frame(Days = df_RUS$Days, y_hat_linear, y_hat_logistic, y_hat_Gompertz)
print(predict %>% filter(Days >= df_RUS_test$Days[1]))
```

```
##     Days y_hat_linear y_hat_logistic y_hat_Gompertz
## 416  416      4393479        4640202        4459640
## 417  417      4403327        4658033        4478334
## 418  418      4413514        4675764        4497011
## 419  419      4424079        4693396        4515671
## 420  420      4435065        4710927        4534314
## 421  421      4446515        4728356        4552938
## 422  422      4458474        4745684        4571545
## 423  423      4470990        4762909        4590133
## 424  424      4484109        4780031        4608702
## 425  425      4497883        4797049        4627252
## 426  426      4512363        4813963        4645782
## 427  427      4527602        4830772        4664292
## 428  428      4543656        4847476        4682782
## 429  429      4560581        4864074        4701251
## 430  430      4578437        4880566        4719699
## 431  431      4597283        4896952        4738126
## 432  432      4617183        4913230        4756531
## 433  433      4638201        4929402        4774914
## 434  434      4660403        4945465        4793275
## 435  435      4683858        4961421        4811613
## 436  436      4708637        4977269        4829928
## 437  437      4734811        4993008        4848219
## 438  438      4762456        5008638        4866487
## 439  439      4791648        5024159        4884731
## 440  440      4822466        5039571        4902950
## 441  441      4854992        5054874        4921145
## 442  442      4889308        5070067        4939315
## 443  443      4925502        5085151        4957460
## 444  444      4963660        5100124        4975579
## 445  445      5003873        5114988        4993672
## 446  446      5046235        5129742        5011740
```

3.1.5 test MSE & R squared

각 모델을 평가하기 위해 먼저, 일일 확진자 수를 얼마나 잘 예측하는지 test MSE를 계산하였다.

```
getMSE <- function(y, yhat){
  MSE <- sum((y-yhat)^2)/length(y)
  return(MSE)
}
MSE_test_linear_daily <- getMSE(y = df_RUS_test$Difference,
                                (y_hat_linear - c(0, y_hat_linear[-length(y_hat_linear)]))[df_RUS_test$Da
MSE_test_logistic_daily <- getMSE(y = df_RUS_test$Difference,
                                (y_hat_logistic - c(0, y_hat_logistic[-length(y_hat_logistic)]))[df_RUS_te
MSE_test_Gompertz_daily <- getMSE(y = df_RUS_test$Difference,
                                (y_hat_Gompertz - c(0, y_hat_Gompertz[-length(y_hat_Gompertz)]))[df_RUS_te
cat("test MSE of linear model :", MSE_test_linear_daily %>% format(scientific = TRUE), "\n")
```

```
## test MSE of linear model : 3.004062e+08
```

```
cat("test MSE of logistic model :", MSE_test_logistic_daily %>% format(scientific = TRUE), "\n")
```

```
## test MSE of logistic model : 7.194612e+07
```

```
cat("test MSE of Gompertz model :", MSE_test_Gompertz_daily %>% format(scientific = TRUE), "\n")
```

```
## test MSE of Gompertz model : 1.105573e+08
```

일일 확진자 수를 예측하는 것은 logistic model이 가장 탁월하였다.

한편, 누적 확진자 수를 얼마나 잘 예측하는지를 측정하기 위해 R squared 값을 계산하였다. 각 모델의 R squared 값은 다음과 같다.

```
getRsq <- function(y, yhat){
  Rsq <- 1 - (sum((y-yhat)^2) / sum((y-mean(y))^2))
  return(Rsq)
}
data.frame(linear = summary(model.linear)$adj.r.squared,
           logistic = getRsq(df_RUS$Cases,y_hat_logistic),
           Gompertz = getRsq(df_RUS$Cases,y_hat_Gompertz)
           )
```

```
##      linear  logistic  Gompertz
## 1 0.9998575 0.9898478 0.9886231
```

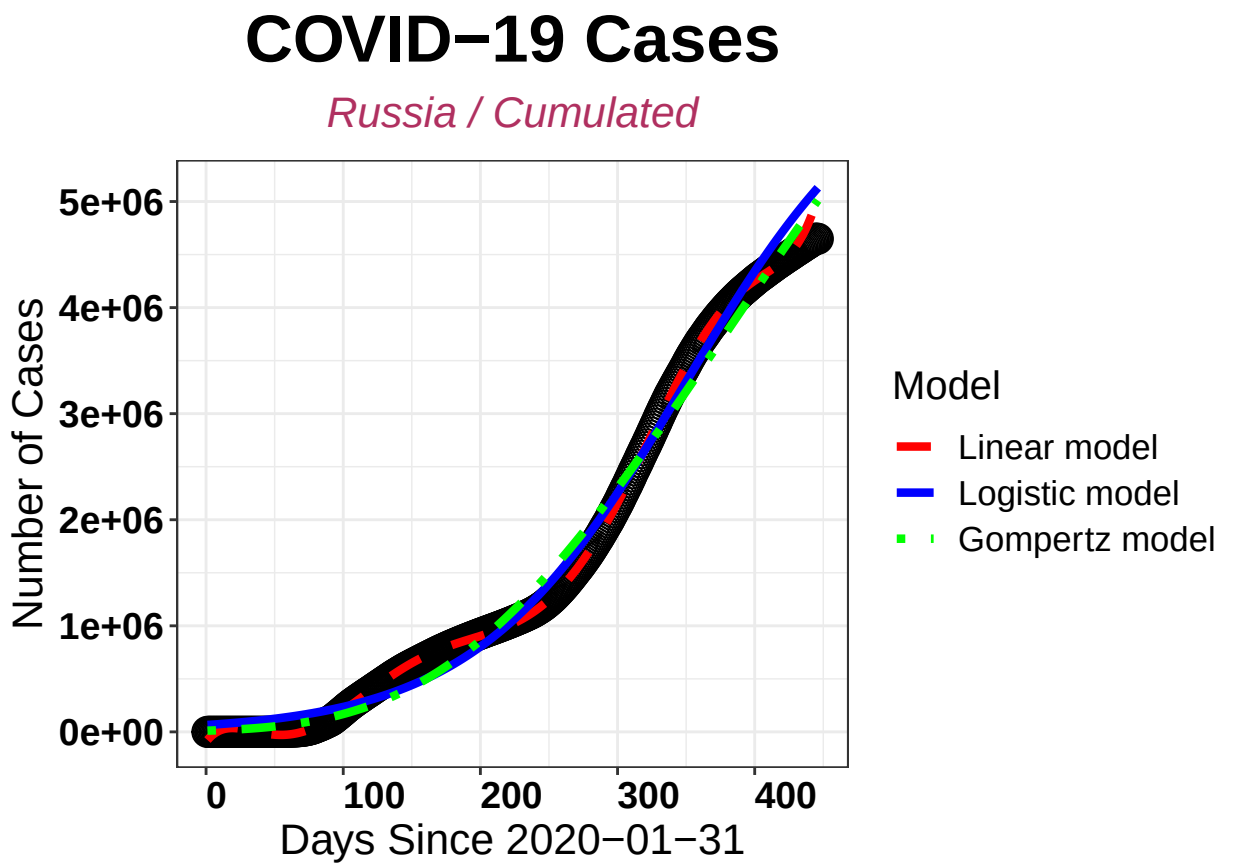
누적 확진자 수 분포를 가장 잘 적합한 모형은 polynomial model 이었다.

3.1.6 Visualization

세 모형을 시각화한 결과는 다음과 같다.

```
df_fitted <- data.frame(Days = df_RUS$Days,
                        y_hat_linear,
                        y_hat_logistic,
                        y_hat_Gompertz)

df_predict <- data.frame(x = rep(df_fitted$Days,3),
                         yhat_cases = c(y_hat_linear, y_hat_logistic, y_hat_Gompertz),
                         yhat_difference = c(y_hat_linear - c(0, y_hat_linear[-length(y_hat_linear)]),
                                             y_hat_logistic - c(0, y_hat_logistic[-length(y_hat_logisti
                                             y_hat_Gompertz - c(0, y_hat_Gompertz[-length(y_hat_Gompert
                         type = rep(c("Linear model", "Logistic model", "Gompertz model"),
                                    each = nrow(df_fitted)))

df_predict$type <- factor(df_predict$type,
                          levels = c("Linear model", "Logistic model", "Gompertz model"))

t0 <- df_RUS$Date[1]
model_labels <- c("Linear model", "Logistic model", "Gompertz model")
models <- list(model.linear, model.logistic, model.Gompertz)

col_list <- c("red", "blue", "green")
shape_list <- c("Linear model"="dashed", "Logistic model"="solid", "Gompertz model"="dotdash")

p_1 <- ggplot(data=df_RUS, aes(x = Days, y = Cases)) +
  geom_point(color='black', shape = 1, size=5) +
  theme_bw() +
  labs(title = paste0("COVID-19 Cases"),
       subtitle = paste0("Russia", " / ", "Cumulated"),
       x = paste0('Days Since ', as.character(t0)),
       y = 'Number of Cases') +
  geom_line(data = df_predict,
            aes(x = x,y = yhat_cases, colour = type, linetype = type), size=1.5)+
  scale_color_manual(name = "Model",
                     labels = model_labels,
                     values = col_list) +
  scale_linetype_manual(name = "Model",
```

```
                        labels = model_labels,
                        values = shape_list) +
  theme(plot.title=element_text(size=25, hjust=0.5, face="bold", colour="black", vjust=2),
        plot.subtitle=element_text(size=16, hjust=0.5, face="italic", color="maroon", vjust=2),
        axis.text=element_text(size=14, face = "bold", colour = "black"),
        axis.text.x = element_text(size = 14, hjust = 0),
        axis.title=element_text(size=16, colour = "black"),
        legend.title = element_text(size = 15),
        legend.text = element_text(size = 13))
p_1
```



누적 확진자 수를 linear model이 가장 잘 적합하였음을 알 수있다. 나머지 두 비선형 모형은 test set에 대해 확진자 수를 과대 추정하는 경향을 보였다.

3.2 Daily Cases with vaccination

검정하고자 하는 가설은 다음과 같다.

H_0 : 백신은 효과가 없다.; H_1 : 백신은 효과가 있다.

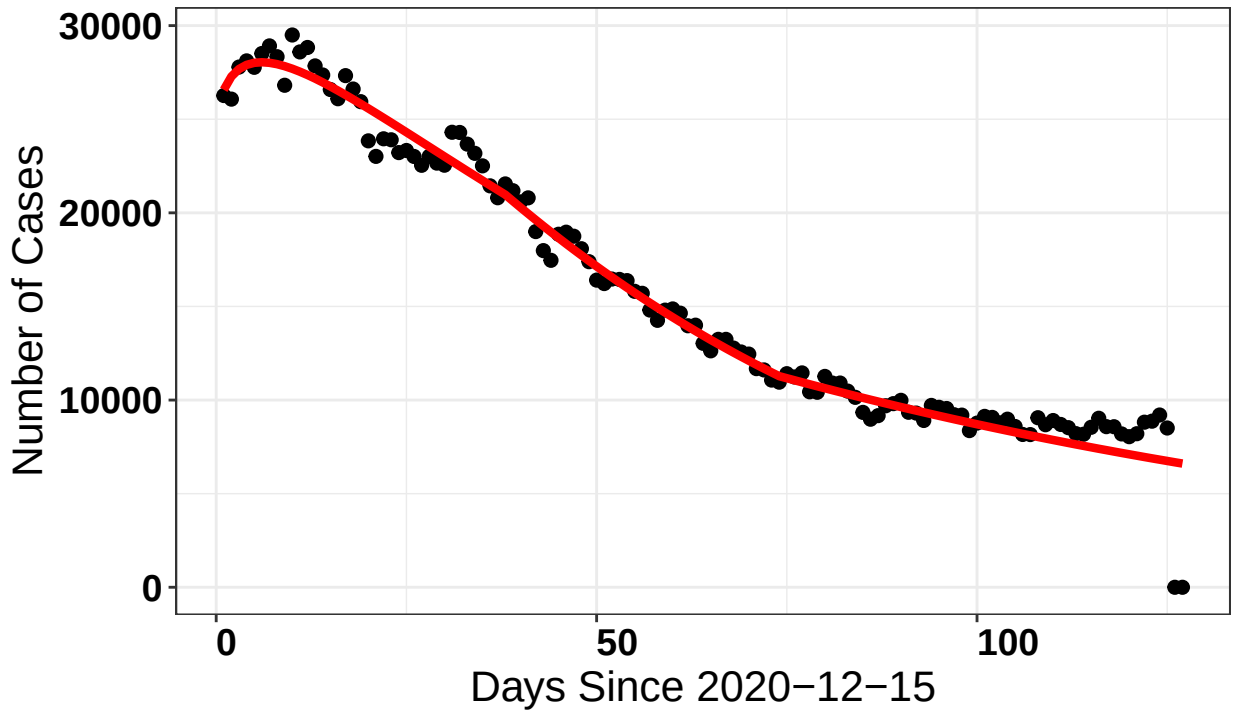
이를 검정하기 위하여 Daily Cases를 Days 변수만으로 적합하여 reduced model을 만들고, Days와 people_vaccinated

변수 모두를 적합한 full model을 비교하여 partial F test를 시행하기로 하였다.

3.2.1 Daily Cases vs Days

reduced model은 다음과 같이 적합되었다.

```
fit_1 <- glm(Difference ~ log(1+Days) + Days, data = vacc_RUS_train, family = poisson)
seg_fit_1 <- segmented(fit_1, seg.Z = ~ Days,
                       npsi = 2, control = seg.control(it.max = 10000, n.boot = 50))



vacc_predict = data.frame(x = vacc_RUS$Days,
                          yhat_cases = exp(predict(seg_fit_1, vacc_RUS)))

t1 <- vacc_RUS$Date[1]
p_2 <- ggplot(data=vacc_RUS, aes(x = Days, y = Difference)) +
  geom_point(color='black', size=2) +
  theme_bw() +
  labs(title = paste0("COVID-19 Cases"),
       subtitle = paste0("Russia", " / ", "daily"),
       x = paste0('Days Since ', as.character(t1)),
       y = 'Number of Cases') +
  geom_line(data = vacc_predict,
            aes(x = x, y = yhat_cases), size=1.5, colour='red')+
  theme(plot.title=element_text(size=25, hjust=0.5, face="bold", colour="black", vjust=2),
        plot.subtitle=element_text(size=16, hjust=0.5, face="italic", color="maroon", vjust=2),
        axis.text=element_text(size=14, face = "bold", colour = "black"),
        axis.text.x = element_text(size = 14, hjust = 0),
        axis.title=element_text(size=16, colour = "black"),
        legend.title = element_text(size = 15),
        legend.text = element_text(size = 13))
p_2
```




```
summary(seg_fit_1)
```

```
##
##  ***Regression Model with Segmented Relationship(s)***
##
## Call:
## segmented.glm(obj = fit_1, seg.Z = ~Days, npsi = 2, control = seg.control(it.max = 10000,
##     n.boot = 50))
##
## Estimated Break-Point(s):
##              Est. St.Err
## psi1.Days 38.001  0.780
## psi2.Days 74.001  0.613
##
## Meaningful coefficients of the linear terms:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)   10.1321833  0.0064180 1578.70   <2e-16 ***
## log(1 + Days)  0.1008705  0.0039835   25.32   <2e-16 ***
## Days          -0.0144965  0.0002824  -51.33   <2e-16 ***
```

```
## U1.Days       -0.0045706  0.0002511  -18.20       NA
## U2.Days        0.0079695  0.0003393   23.49       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null     deviance: 231522.5  on 95  degrees of freedom
## Residual deviance:   2673.5  on 89  degrees of freedom
## AIC: 3798.5
##
## Convergence attained in 7 iter. (rel. change 4.2858e-06)
```

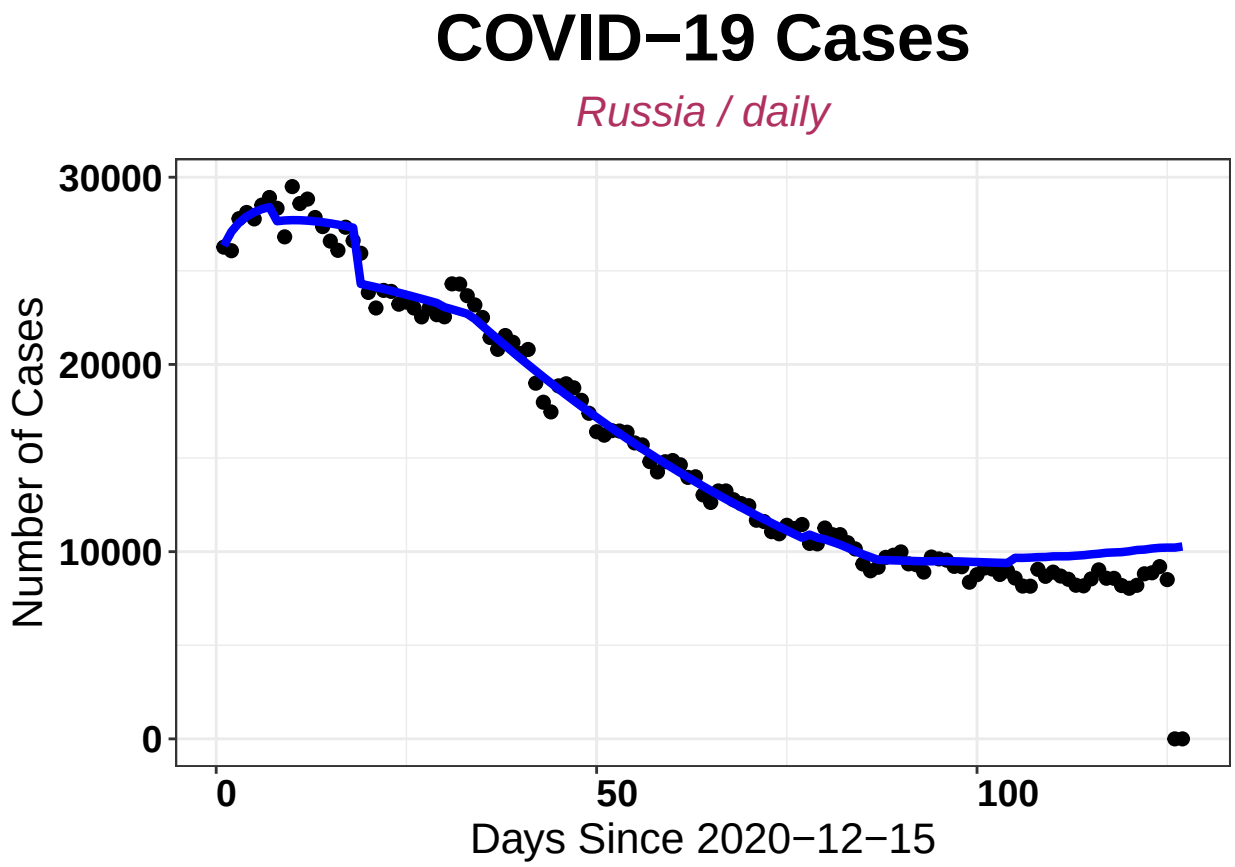
test set을 과소추정 하였지만, 비교적 잘 적합되었다고 판단하였다.

3.2.2 Daily Cases vs Days + # of People Vaccinated

full model은 다음과 같이 적합되었다.

```
fit_2 <- glm(Difference~ log(1+Days) + Days + log(1+people_vaccinated) + people_vaccinated,
             data = vacc_RUS_train, family = poisson)
seg_fit_2 <- segmented(fit_2, seg.Z = ~ Days,
                       npsi = 2, control = seg.control(it.max = 10000, n.boot = 50))



vacc_predict = data.frame(x = vacc_RUS$Days,
                          yhat_cases = exp(predict(seg_fit_2, vacc_RUS)))

t1 <- vacc_RUS$Date[1]
p_3 <- ggplot(data=vacc_RUS, aes(x = Days, y = Difference)) +
  geom_point(color='black', size=2) +
  theme_bw() +
  labs(title = paste0("COVID-19 Cases"),
       subtitle = paste0("Russia", " / ", "daily"),
       x = paste0('Days Since ', as.character(t1)),
       y = 'Number of Cases') +
  geom_line(data = vacc_predict,
            aes(x = x, y = yhat_cases), size=1.5, colour='blue')+
  theme(plot.title=element_text(size=25, hjust=0.5, face="bold", colour="black", vjust=2),
        plot.subtitle=element_text(size=16, hjust=0.5, face="italic", color="maroon", vjust=2),
        axis.text=element_text(size=14, face = "bold", colour = "black"),
```

```
        axis.text.x = element_text(size = 14, hjust = 0),
        axis.title=element_text(size=16, colour = "black"),
        legend.title = element_text(size = 15),
        legend.text = element_text(size = 13))
p_3
```




```
summary(seg_fit_2)
```

```
##
##  ***Regression Model with Segmented Relationship(s)***
##
## Call:
## segmented.glm(obj = fit_2, seg.Z = ~Days, npsi = 2, control = seg.control(it.max = 10000,
##     n.boot = 50))
##
## Estimated Break-Point(s):
##               Est. St.Err
## psi1.Days 33.346  0.335
## psi2.Days 87.000  0.429
##
```

```
## Meaningful coefficients of the linear terms:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  1.064e+01  2.212e-02  481.13   <2e-16 ***
## log(1 + Days)                8.867e-02  4.554e-03   19.47   <2e-16 ***
## Days                        -7.889e-03  4.855e-04  -16.25   <2e-16 ***
## log(1 + people_vaccinated) -5.047e-02  2.034e-03  -24.81   <2e-16 ***
## people_vaccinated           3.320e-08  1.515e-09   21.92   <2e-16 ***
## U1.Days                     -1.087e-02  4.274e-04  -25.44       NA
## U2.Days                      1.643e-02  1.132e-03   14.52       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null     deviance: 231522.5  on 95  degrees of freedom
## Residual deviance:   1941.8  on 87  degrees of freedom
## AIC: 3070.8
##
## Convergence attained in 6 iter. (rel. change 7.137e-07)
```

test set을 과대추정하였지만, 비교적 잘 적합하였다고 판단하였다.

3.2.3 Partial F test

```
anova(seg_fit_1, seg_fit_2)
```

```
## Analysis of Deviance Table
##
## Model 1: Difference ~ log(1 + Days) + Days + U1.Days + U2.Days + psi1.Days +
##     psi2.Days
## Model 2: Difference ~ log(1 + Days) + Days + log(1 + people_vaccinated) +
##     people_vaccinated + U1.Days + U2.Days + psi1.Days + psi2.Days
##   Resid. Df Resid. Dev Df Deviance
## 1        89     2673.5
## 2        87     1941.8  2   731.73
```

```
F0 = (731.73/2)/(1941.8/87)
F0 > qf(0.05, 2, 87, lower.tail=FALSE)
```

```
## [1] TRUE
```

F statistics 값이 기각역에 포함되어, H_0를 기각할 수 있었다. 따라서 백신은 일일 확진자 수에 영향을 미친다고 할 수 있다.

그러나 모델 summary 에서 알 수 있듯이, people_vaccinated의 계수가 매우 작지만 양수이므로, 백신이 코로나 확진자 수 감소에 영향을 미친다고 해석하기는 어렵다고 판단한다.

4. Conclusion

코로나 누적 확진자 수를 여러 모형으로 적합하였다. 그 중 polynomial model이 비선형 모형들보다 누적 확진자 수를 잘 설명하였다.

일일 확진자 수를 적합 할 때, 백신 접종자수를 regressor로 추가했을 때 fitting은 매우 개선되었으나, 백신이 코로나 확진자 수를 뚜렷이 감소시킨다고는 판단할 수 없었다.