

13주차 과제

서울대학교 통계학과 2017-11362 박건도

2021년 05월 30일

```
data <- read.csv("owid_covid_210518.csv")
data <- data %>% filter(location == "Romania") %>% select(-location)
first_case <- data[1, "date"]
data <- data %>% # scaling
  mutate(days = as.Date(date) - as.Date(first_case) + 1,
         days = as.numeric(days),
         logdays = log(days),
         new_cases_per_million = new_cases_per_million * 1000,
         new_deaths_per_million = new_deaths_per_million * 1000,
         icu_patients_per_million = icu_patients_per_million * 1000,
         people_vaccinated_per_hundred = people_vaccinated_per_hundred)
```

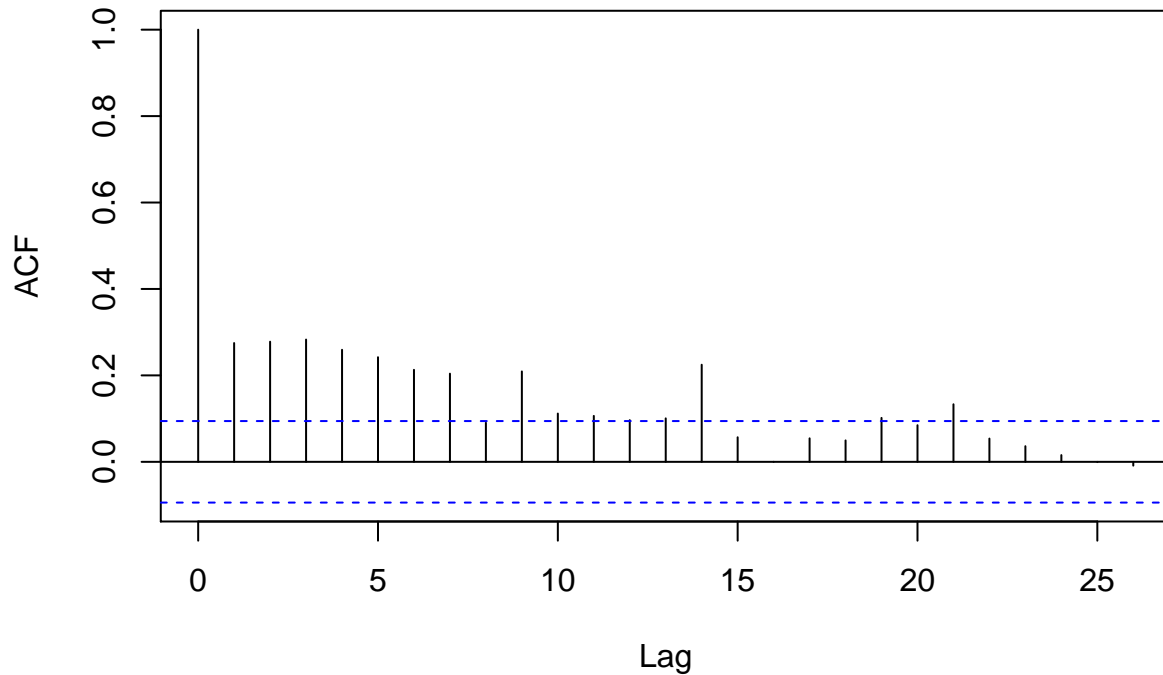
1. 일일 신규 확진자 수

1) 백신 효과를 나타내는 모수의 정의 및 추정

```
timeseries <- ts(data[, "new_cases_per_million"])
regressors <- data[, c("people_vaccinated_per_hundred",
                      "days", "logdays")]
timeseries_train <- window(timeseries, end = 432)
regressors_train <- window(ts(regressors), end = 432)
timeseries_test <- window(timeseries, start = 433, end = 447)
regressors_test <- window(ts(regressors), start = 433, end = 447)

vaccine_fit <- tsglm(timeseries_train,
                    model = list(past_obs = c(1,7), past_mean = c(1)),
                    link = "log", distr = "poisson",
                    xreg = regressors_train)
acf(residuals(vaccine_fit))
```

Series residuals(vaccine_fit)



ACF 그래프를 보면 0을 제외하고는 전반적으로 0.2 미만의 autocorrelation을 갖는다. 위 적합 결과의 vaccinated_people_per_hundred의 계수가 바로 백신 효과를 나타내는 모수가 된다.

2) 백신 효과에 대한 유의성 검정

```
summary(vaccine_fit)
```

Call:

```
tsglm(ts = timeseries_train, model = list(past_obs = c(1, 7),
    past_mean = c(1)), xreg = regressors_train, link = "log",
    distr = "poisson")
```

Coefficients:

	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	-0.34538	1.02e-02	-0.36528	-0.32547
beta_1	0.61492	6.10e-04	0.61372	0.61611
beta_7	0.72984	4.57e-04	0.72894	0.73074
alpha_1	-0.40968	6.16e-04	-0.41089	-0.40848
people_vaccinated_per_hundred	-0.00339	9.70e-05	-0.00358	-0.00320
days	-0.00143	1.36e-05	-0.00146	-0.00141
logdays	0.28187	2.59e-03	0.27678	0.28695

Standard errors and confidence intervals (level = 95 %) obtained

by normal approximation.

Link function: log

Distribution family: poisson

Number of coefficients: 7

Log-likelihood: -996306.8

AIC: 1992628

BIC: 1992656

QIC: 1992628

people_vaccinated_per_hundred의 계수를 보면 -0.00339로, Confidence Interval 또한 -0.003 근처이다. 따라서 유의하게 일일 확진자수를 감소시킨다고 할 수 있다.

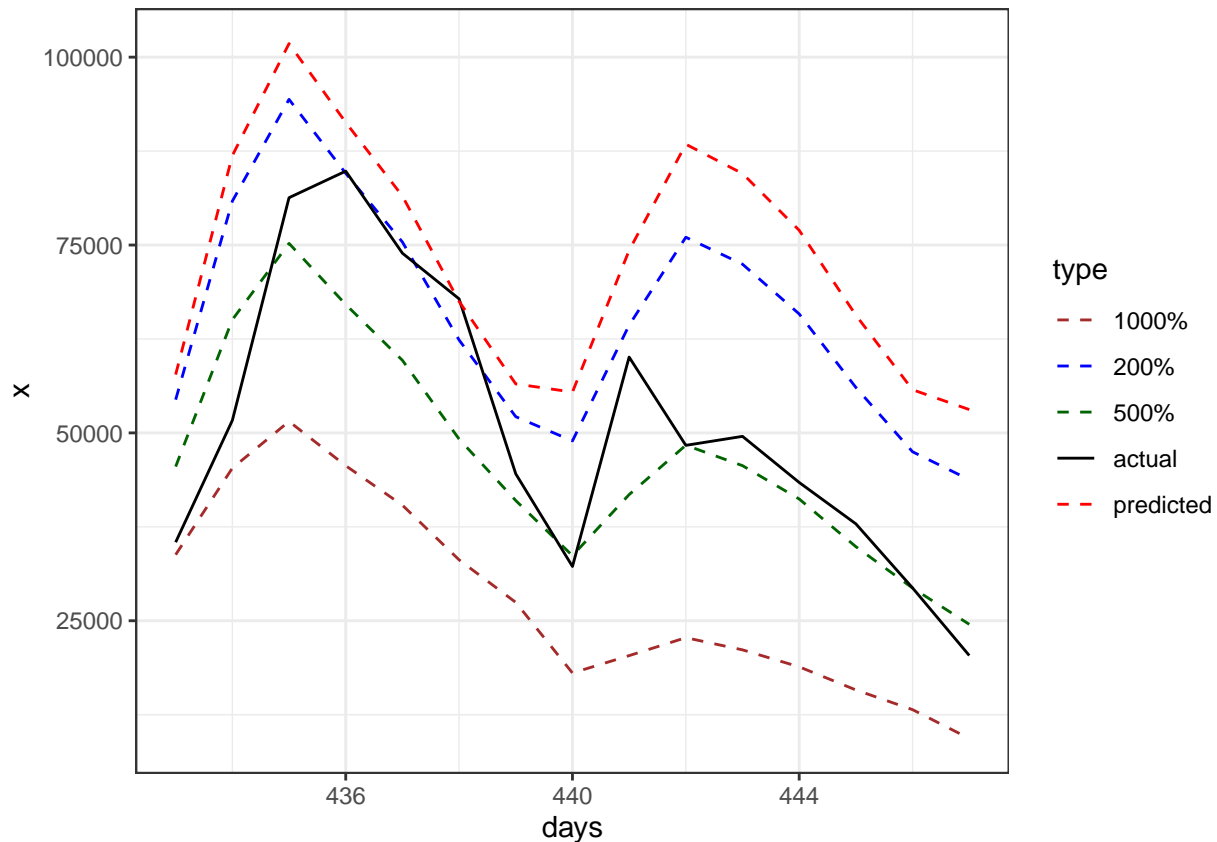
3) 추정된 백신 효과를 기반으로 한 향후 신규 확진자 수의 추정

```
a = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test)
regressors_test_2 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 2 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
b = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_2)
regressors_test_3 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 5 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
c = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_3)
regressors_test_4 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 10 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
d = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_4)

result <- data.frame(
  days = rep(433:447, 5),
  x = c(timeseries_test, a$pred, b$pred, c$pred, d$pred),
  type = rep(c("actual", "predicted", "200%", "500%", "1000%"), each = 15))

ggplot(data = result, aes(x = days, y = x, color = type, linetype = type)) +
  geom_line() +
```

```
scale_linetype_manual(values = c(rep("dashed", 3), "solid", "dashed"))+
scale_color_manual(values = c("brown", "blue", "darkgreen", "black", "red"))+
theme_bw()
```



백신 접종률을 기존의 predicted에서 2배, 5배, 10배 늘린 결과, 예측된 신규 확진자 수는 점점 감소함을 알 수 있다.

2. 일일 사망자 수

1) 백신 효과를 나타내는 모수의 정의 및 추정

```
timeseries <- ts(data[, "new_deaths_per_million"])
regressors <- data[, c("people_vaccinated_per_hundred",
                       "new_cases_per_million",
                       "days", "logdays")]
timeseries_train <- window(timeseries, end = 432)
regressors_train <- window(ts(regressors), end = 432)
timeseries_test <- window(timeseries, start = 433, end = 447)
regressors_test <- window(ts(regressors), start = 433, end = 447)

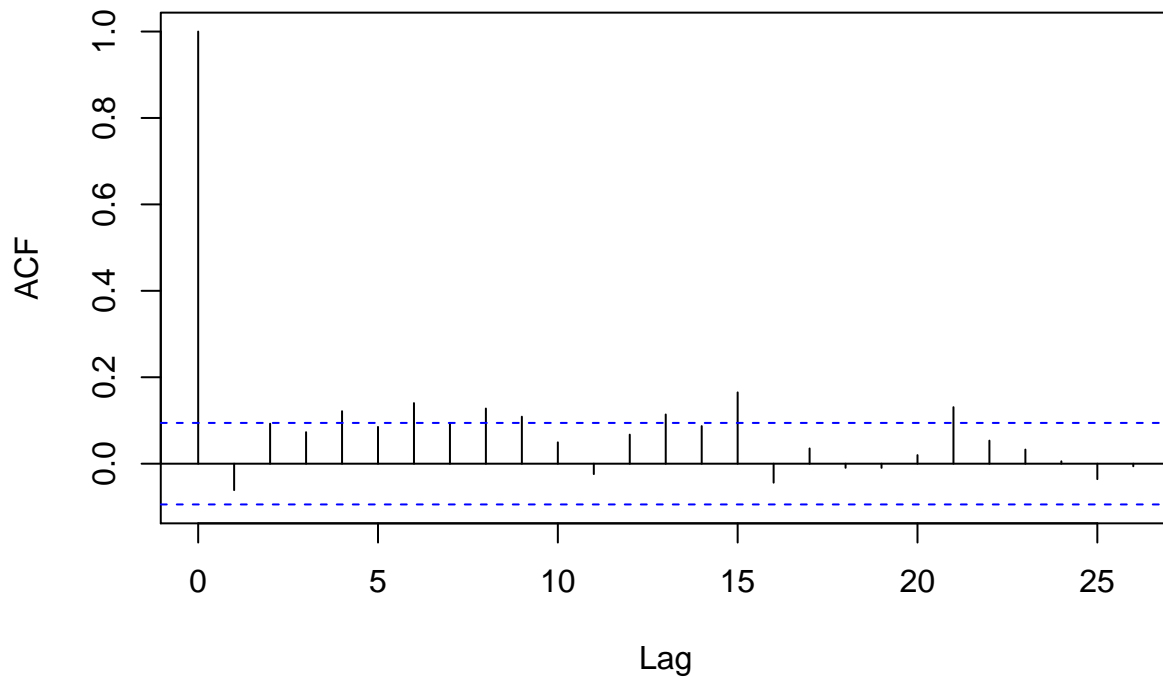
vaccine_fit <- tsglm(timeseries_train,
```

```

model = list(past_obs = c(1,7), past_mean = c(1)),
link = "log", distr = "poisson",
xreg = regressors_train)
acf(residuals(vaccine_fit))

```

Series residuals(vaccine_fit)



ACF 그래프를 보면 0을 제외하고는 전반적으로 0.1 근처의 autocorrelation을 갖는다. 위 적합 결과의 vaccinated_people_per_hundred의 계수가 바로 백신 효과를 나타내는 모수가 된다.

2) 백신 효과에 대한 유의성 검정

```
summary(vaccine_fit)
```

Call:

```

tsglm(ts = timeseries_train, model = list(past_obs = c(1, 7),
past_mean = c(1)), xreg = regressors_train, link = "log",
distr = "poisson")

```

Coefficients:

	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	3.46e+00	3.62e-02	3.39e+00	3.53e+00
beta_1	3.22e-01	3.21e-03	3.16e-01	3.28e-01
beta_7	4.13e-01	2.62e-03	4.08e-01	4.18e-01

alpha_1	-2.37e-01	4.37e-03	-2.45e-01	-2.28e-01
people_vaccinated_per_hundred	2.84e-05	4.16e-04	-7.86e-04	8.43e-04
new_cases_per_million	1.51e-06	1.19e-08	1.49e-06	1.53e-06
days	1.94e-03	5.07e-05	1.84e-03	2.04e-03
logdays	-1.75e-02	7.93e-03	-3.31e-02	-2.00e-03

Standard errors and confidence intervals (level = 95 %) obtained by normal approximation.

Link function: log

Distribution family: poisson

Number of coefficients: 8

Log-likelihood: -42903.3

AIC: 85822.59

BIC: 85855.14

QIC: 85822.51

people_vaccinated_per_hundred의 계수를 보면 $-2.84e-05$ 로, Confidence Interval은 $[-7.86e-04, 8.43e-04]$ 로 0을 포함한다. 따라서 백신 접종률은 사망자 수에 유의미한 영향을 끼치지 않는다고 할 수 있다.

3) 추정된 백신 효과를 기반으로 한 향후 사망자 수의 추정

```
a = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test)
regressors_test_2 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 2 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
b = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_2)
regressors_test_3 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 5 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
c = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_3)
regressors_test_4 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 10 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
d = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_4)

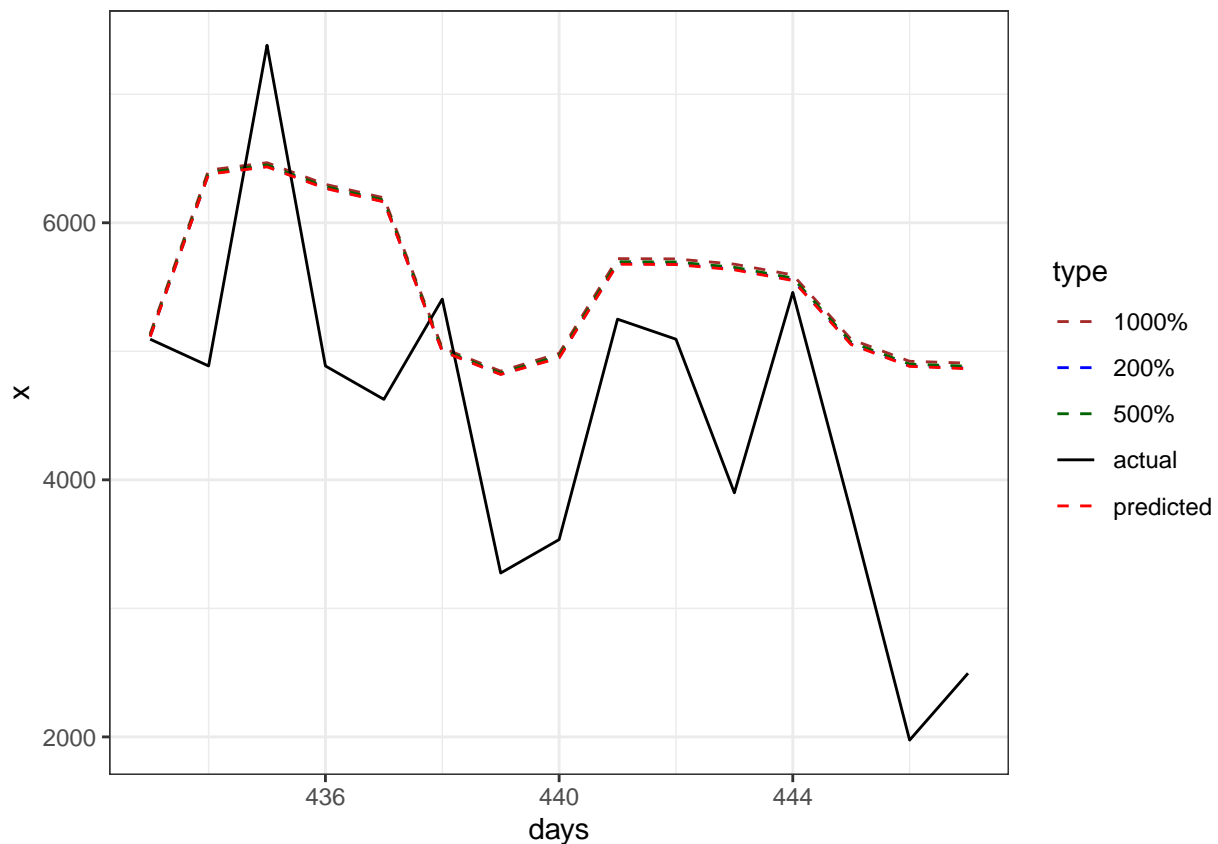
result <- data.frame(
```

```

days = rep(433:447, 5),
x = c(timeseries_test, a$pred, b$pred, c$pred, d$pred),
type = rep(c("actual", "predicted", "200%", "500%", "1000%"), each = 15))

ggplot(data = result, aes(x = days, y = x, color = type, linetype = type)) +
  geom_line()+
  scale_linetype_manual(values = c(rep("dashed", 3), "solid", "dashed"))+
  scale_color_manual(values = c("brown", "blue", "darkgreen", "black", "red"))+
  theme_bw()

```



애초에 백신 접종률이 유의미한 영향을 끼치지 않으므로, 백신 접종률을 늘렸을 때 큰 변화가 생기지 않는다.

3. 일일 ICU 환자 수

1) 백신 효과를 나타내는 모수의 정의 및 추정

```

timeseries <- ts(data[, "icu_patients_per_million"])
regressors <- data[, c("people_vaccinated_per_hundred",
                       "new_cases_per_million",
                       "days", "logdays")]

```

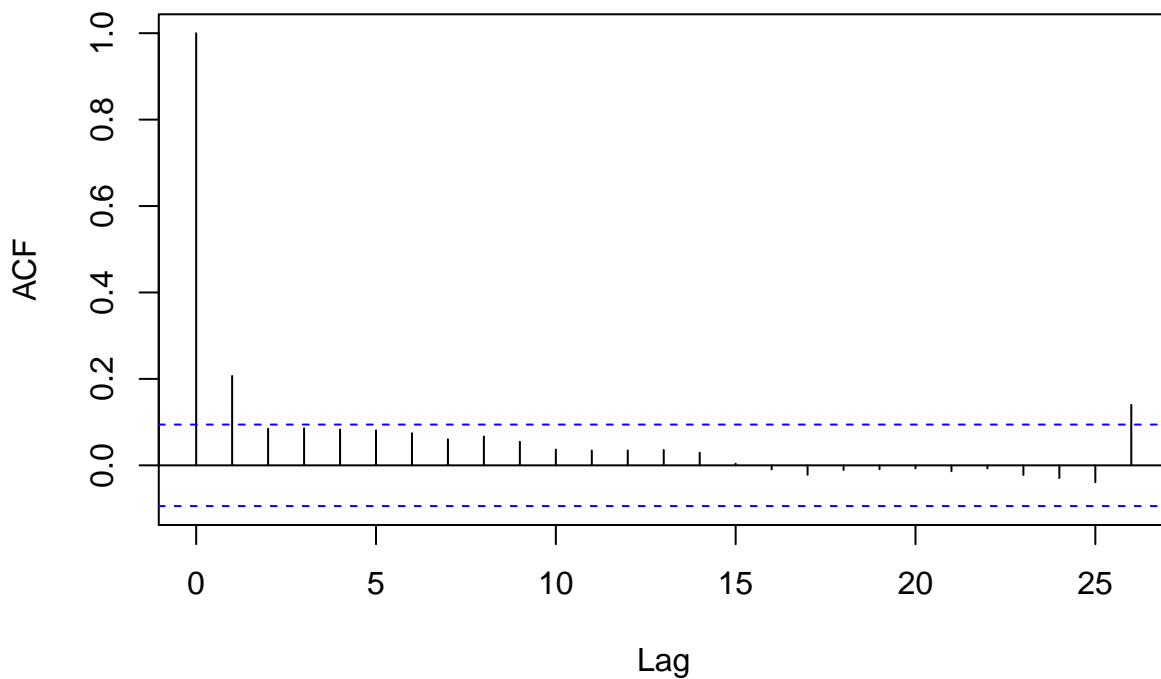
```

timeseries_train <- window(timeseries, end = 432)
regressors_train <- window(ts(regressors), end = 432)
timeseries_test <- window(timeseries, start = 433, end = 447)
regressors_test <- window(ts(regressors), start = 433, end = 447)

vaccine_fit <- tsglm(timeseries_train,
                     model = list(past_obs = c(1,7), past_mean = c(1)),
                     link = "log", distr = "poisson",
                     xreg = regressors_train)
acf(residuals(vaccine_fit))

```

Series residuals(vaccine_fit)



ACF 그래프를 보면 0을 제외하고는 전반적으로 0.1 미만의 autocorrelation을 갖는다. 위 적합 결과의 vaccinated_people_per_hundred의 계수가 바로 백신 효과를 나타내는 모수가 된다.

2) 백신 효과에 대한 유의성 검정

```
summary(vaccine_fit)
```

Call:

```

tsglm(ts = timeseries_train, model = list(past_obs = c(1, 7),
      past_mean = c(1)), xreg = regressors_train, link = "log",
      distr = "poisson")

```


Coefficients:

	Estimate	Std.Error	CI(lower)	CI(upper)
(Intercept)	6.07e-01	2.31e-03	6.02e-01	6.11e-01
beta_1	1.50e-02	7.64e-05	1.49e-02	1.52e-02
beta_7	5.55e-04	8.27e-05	3.93e-04	7.17e-04
alpha_1	8.93e-01	4.21e-04	8.92e-01	8.94e-01
people_vaccinated_per_hundred	-1.27e-04	1.29e-05	-1.52e-04	-1.02e-04
new_cases_per_million	1.75e-07	6.51e-10	1.74e-07	1.76e-07
days	1.34e-04	1.46e-06	1.31e-04	1.37e-04
logdays	5.45e-02	3.30e-04	5.38e-02	5.51e-02

Standard errors and confidence intervals (level = 95 %) obtained
by normal approximation.

Link function: log

Distribution family: poisson

Number of coefficients: 8

Log-likelihood: -555847.5

AIC: 1111711

BIC: 1111744

QIC: 1111711

people_vaccinated_per_hundred의 계수를 보면 $-1.27e-04$ 로, Confidence Interval 또한 $-1.02e-04$ 를 넘지 않는. 따라서 유의하게 일일 ICU 환자 수를 감소시킨다고 할 수 있다.

3) 추정된 백신 효과를 기반으로 한 향후 일일 ICU 환자 수의 추정

```
a = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test)
regressors_test_2 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 2 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
b = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_2)
regressors_test_3 <- regressors %>%
  mutate(people_vaccinated_per_hundred = 5 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
c = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_3)
regressors_test_4 <- regressors %>%
```

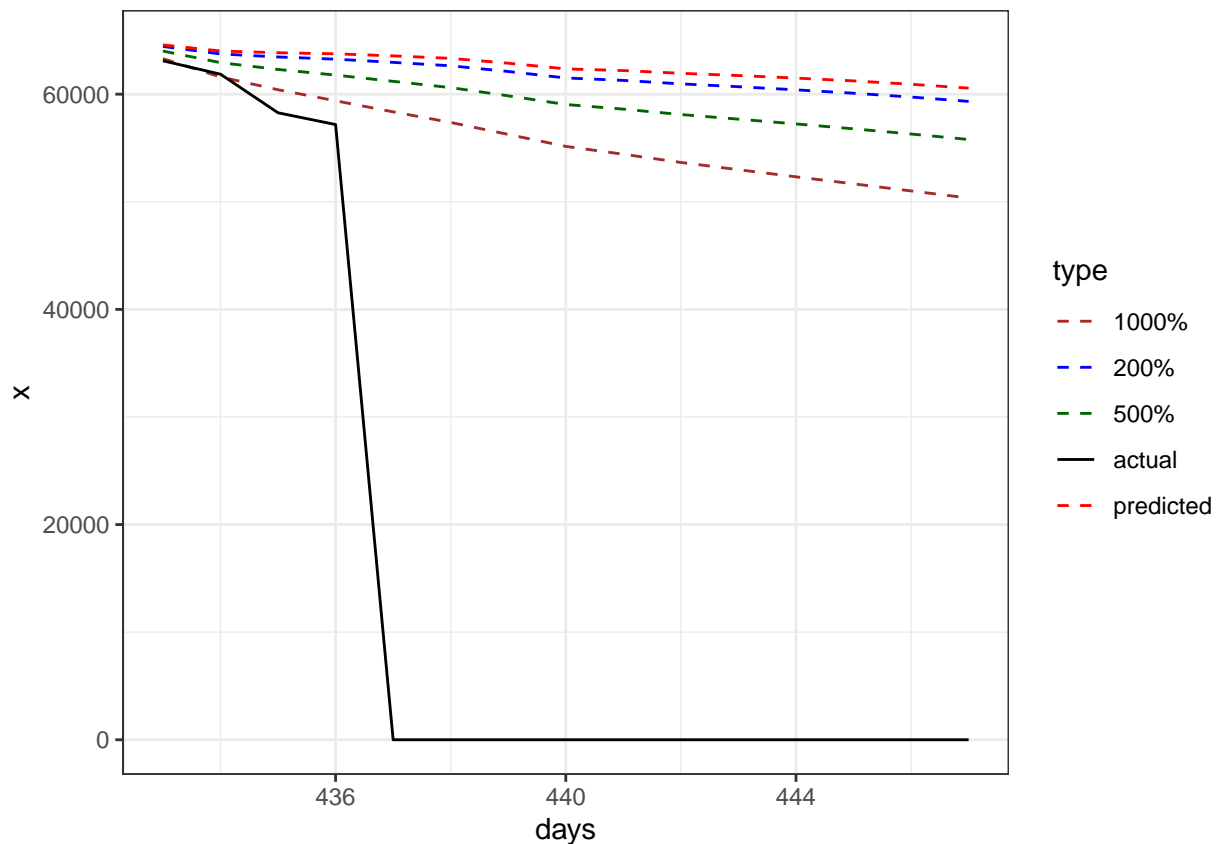
```

mutate(people_vaccinated_per_hundred = 10 * people_vaccinated_per_hundred) %>%
  ts() %>% window(start = 433, end = 447)
d = predict(vaccine_fit, n.ahead = 15, level = 0.9, global = TRUE,
            B = 2000, newxreg = regressors_test_4)

result <- data.frame(
  days = rep(433:447, 5),
  x = c(timeseries_test, a$pred, b$pred, c$pred, d$pred),
  type = rep(c("actual", "predicted", "200%", "500%", "1000%"), each = 15))

ggplot(data = result, aes(x = days, y = x, color = type, linetype = type)) +
  geom_line()+
  scale_linetype_manual(values = c(rep("dashed", 3), "solid", "dashed"))+
  scale_color_manual(values = c("brown", "blue", "darkgreen", "black", "red"))+
  theme_bw()

```



백신 접종률을 기존의 predicted에서 2배, 5배, 10배 늘린 결과, 예측된 신규 확진자 수는 점점 감소함을 알 수 있다.