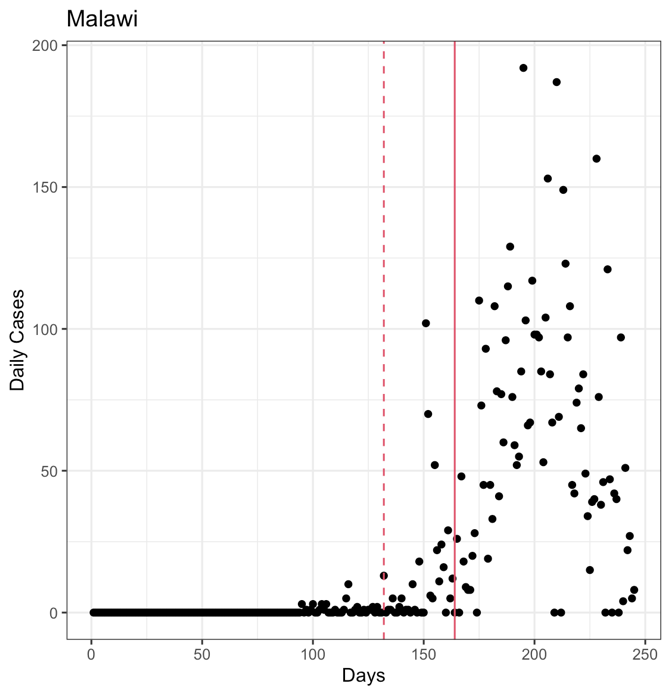
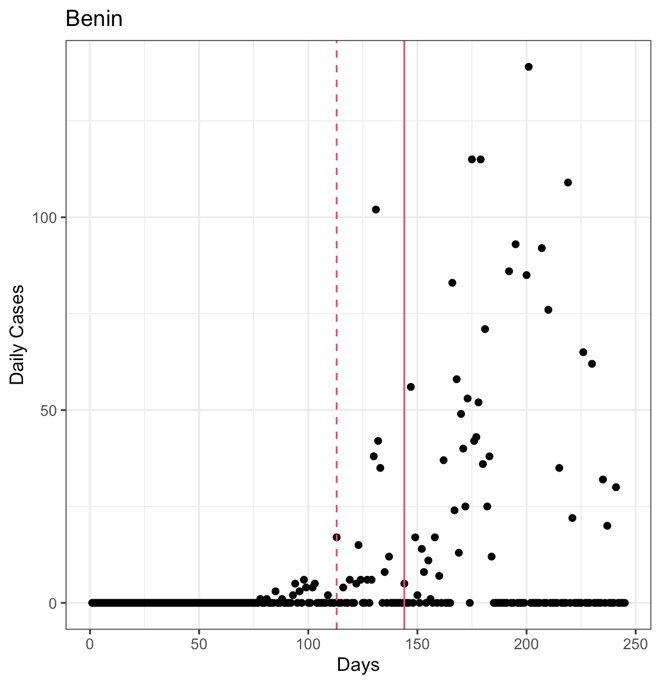
Logistic, Gompertz Model에서 **segment2의 계수b**에 대한 문제점

b 모수 자체에 대해 더 알아보았을 때 2개국(Malawi, Benin)을 제외하고는 모두 2nd Segment에서의 b가 항상 큰 것을 알 수 있었습니다. Logistic, Gompertz Model 모두 Benin, Malawi에서만 b1 > b2로 나타났습니다. plot에서 확인했더니 두 국가 모두 Segment가 1로 확인되는데 2개의 segment로 잘못 적합된 국가로 보입니다.



**(점선은 startingpoint / 실선은 breakpoint)**

Benin, Malawi 두 국가를 제외한다면 연속성을 보장하지 않은 segmented growth curve를 사용했음에도 항상 b1 < b2인 현상이 이상했습니다. 아무래도 b2가 shift의 의미를 담고 있는 만큼 현재 모델에서 b2를 해석함에 있어 b1 또는 두 번째 segment라는 점에 크게 영향을 받는 것 같습니다.

즉, 보정을 해줘야 하는데, logistic의 경우 boxplot에서 segment끼리 scale에 큰 차이가 없어 단순 뺄셈(b2-b1)도 가능하다고 생각합니다. 하지만 Gompertz의 경우 boxplot에서 Q1, Q3 차이도 컸지만 값이 엄청나게 큰 계수들이 여러 개 존재했고 실제로 표준편차를 구해보면 segment1에서는 20, segment2에서는 32000 정도로 큰 차이가 있었습니다. 따라서 Gompertz의 경우 어떻게 보정을 할 것인지에 대해 논의가 필요합니다.

Chart, box and whisker chart

Description automatically generated

표준편차 : 각각 2.7, 5.6

Chart, histogram

Description automatically generated

표준편차 : 각각 20.2, 319506.7

[R-Code]

#rm(list=ls())

library(dplyr)

library(ggplot2)

library(nls2)

# Analysis Period( ~ 8/31)

max\_date <- set\_date("2020/8/31")

#### Data import & Data Preprocessing & Peak Detection Analysis ####

# data import

# df <- read.csv("https://opendata.ecdc.europa.eu/covid19/casedistribution/csv", na.strings = "", fileEncoding = "UTF-8-BOM")

# write.csv(df, "covid\_data.csv")

df <- read.csv("covid\_data.csv")

coef <- read.csv("coef\_result.csv")

setwd("~/Documents/GitHub/COVID-19/201218\_coefficient\_b")

getwd()

# generate vector of country names

country <- as.character(unique(df$countriesAndTerritories))

country <- country[-which(country=="Cases\_on\_an\_international\_conveyance\_Japan")]

# preprocess data (until 20/08/31)

df\_sum = preprocessing\_data()[1:245,]

Malawi <- df\_sum['Malawi']

Malawi\_st <- coef[which(coef$X=="Malawi"),]$startpoint

Malawi\_bp <- coef[which(coef$X=="Malawi"),]$breakpoint

Andorra <- df\_sum['Andorra']

Andorra\_st <- coef[which(coef$X=="Andorra"),]$startpoint

Andorra\_bp <- coef[which(coef$X=="Andorra"),]$breakpoint

Benin <- df\_sum['Benin']

Benin\_st <- coef[which(coef$X=="Benin"),]$startpoint

Benin\_bp <- coef[which(coef$X=="Benin"),]$breakpoint

ggplot() +

geom\_point(aes(x=1:nrow(Malawi), y=malawi$Malawi)) +

geom\_vline(aes(xintercept=Malawi\_st), color=2, linetype=2) +

geom\_vline(aes(xintercept=Malawi\_bp), color=2) +

theme\_bw() +

labs(title="Malawi", x="Days", y="Daily Cases")

ggplot() +

geom\_point(aes(x=1:nrow(Andorra), y=Andorra$Andorra)) +

geom\_vline(aes(xintercept=Andorra\_st), color=2, linetype=2) +

geom\_vline(aes(xintercept=Andorra\_bp), color=2) +

theme\_bw() +

labs(title="Andorra", x="Days", y="Daily Cases")

ggplot() +

geom\_point(aes(x=1:nrow(Benin), y=benin$Benin)) +

geom\_vline(aes(xintercept=Benin\_st), color=2, linetype=2) +

geom\_vline(aes(xintercept=Benin\_bp), color=2) +

theme\_bw() +

labs(title="Benin", x="Days", y="Daily Cases")

coef[which(coef$b1\_Logi - coef$b2\_Logi >0 ),]

coef[which(coef$b1\_Gom - coef$b2\_Gom >0 ),]

# Benin, Malawi, Andorra 제거

coef <- coef[-which(coef$X%in%c("Benin", "Andorra", "Malawi")),]

boxplot(coef$b1\_Logi, coef$b2\_Logi, main="b\_Logistic", xlab="Segment")

boxplot(coef$b1\_Gom, coef$b2\_Gom, main="b\_Gompertz", xlab="Segment")

# outlier : Kuwait(b2\_Gom), Kyrgyzstan(b1\_Gom)

hist(coef$b1\_Logi - coef$b2\_Logi, main="b1 - b2 for Logistic Model")

hist(coef$b1\_Gom - coef$b2\_Gom, main="b1 - b2 for Gompertz Model", breaks = 100)

coef <- coef[-which(coef$X%in%c("Kuwait", "Kyrgyzstan")),]

hist(coef$b1\_Gom - coef$b2\_Gom, main="b1 - b2 for Gompertz Model", breaks = 100)

#### coefficient b와 c와의 관계

ggplot(coef) +

geom\_point(aes(coef$b1\_Logi, coef$c1\_Logi)) +

theme\_bw()

ggplot(coef) +

geom\_point(aes(coef$b2\_Logi- coef$b1\_Logi, coef$c2\_Logi)) +

theme\_bw()

ggplot(coef) +

geom\_point(aes(coef$b1\_Gom, coef$c1\_Gom)) +

theme\_bw()

ggplot(coef) +

geom\_point(aes(coef$b2\_Gom, coef$c2\_Gom)) +

theme\_bw()

ggplot(coef) +

geom\_point(aes(coef$b2\_Gom - coef$b1\_Logi, coef$c2\_Gom)) +

theme\_bw()