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.

Chart, scatter chart

Description automatically generated

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]

pp1 <- ggplot()+

geom\_point(aes(1:length(df\_sum$Malawi), df\_sum$Malawi), alpha=0.5) +

labs(title = "Malawi") +

theme\_bw()

pp2 <- ggplot()+

geom\_point(aes(1:length(df\_sum$Benin), df\_sum$Benin), alpha=0.5) +

labs(title = "Benin") +

theme\_bw()

gridExtra::grid.arrange(pp1, pp2)

setwd("/Users/yeonghyeon/Documents/GitHub/COVID-19/201116\_Permutation\_Test")

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

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

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

# Andorra (b2\_Logi=401) / Kyrgyzstan (b1\_Gom=5915) / Benin (b1\_Gom=3454) / Kuwait (b2\_Gom)

par(mfrow=c(1,2))

boxplot(coef[-c(4,18,81,82),]$b1\_Logi, main="b1\_Logi"); boxplot(coef[-c(4,18,81,82),]$b2\_Logi, main="b2\_Logi")

boxplot(coef[-c(4,18,81,82),]$b1\_Gom, main="b1\_Gom"); boxplot(coef[-c(4,18,81,82),]$b2\_Gom, main="b2\_Gom")

summary(coef[-c(4,18,81,82),]$b1\_Gom)

summary(coef[-c(4,18,81,82),]$b2\_Gom)

var(coef[-c(4,18,81,82),]$b2\_Gom)

sd(coef[-c(4,18,81,82),]$b1\_Logi[!is.na(coef[-c(4,18,81,82),]$b1\_Logi)])

sd(coef[-c(4,18,81,82),]$b2\_Logi[!is.na(coef[-c(4,18,81,82),]$b2\_Gom)])

sd(coef[-c(4,18,81,82),]$b1\_Gom[!is.na(coef[-c(4,18,81,82),]$b1\_Gom)])

sd(coef[-c(4,18,81,82),]$b2\_Gom[!is.na(coef[-c(4,18,81,82),]$b2\_Gom)])