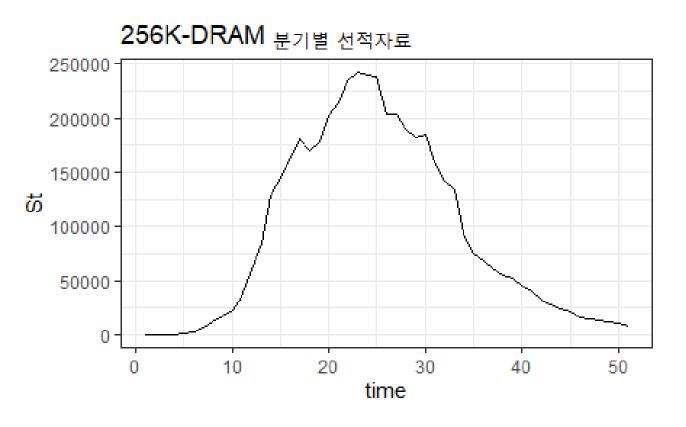


# Contents

- 1. DRAM 총수요 예측
- 2. 영국 COVID-19 & 미국·한국 HIV/AIDS 예측
- 3. 영화 흥행 예측
- 4. R Shiny App



#### 1. 256K-DRAM 분기별 선적자료에 대한 시계열 도표



1982년부터 1995년까지의 DRAM의 분기별 판매 수량이다.

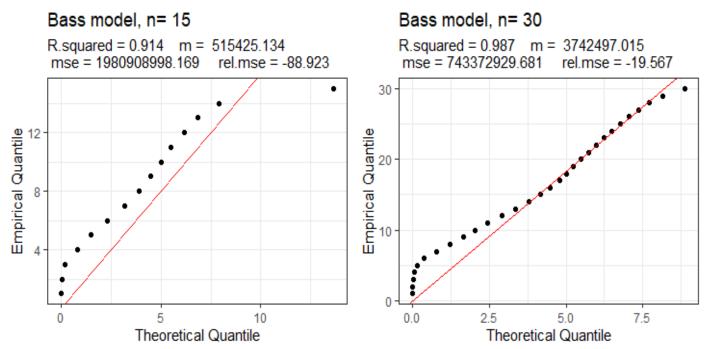
1988년도에 가장 수요가 많았으며 peak 이후로는 계속 감소하는 패턴을 보이고 있다.

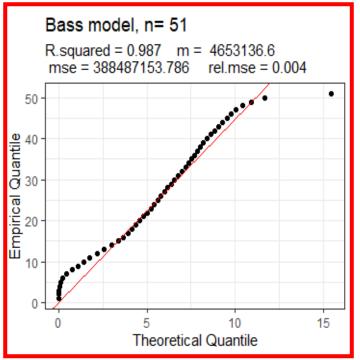
## 2. Bass, Logistic, Gumbel 모델을 이용하여 OLS (m, p, q) 추정

	Bass	Logistic	Gumbel
n	$(\widehat{m},\widehat{p},\widehat{q})$	$(\widehat{m},\widehat{q})$	$(\widehat{m},\widehat{q})$
15	(892463, 0.00067, 0.681)	(881245.6, 0.688)	(5890455, 0.1487)
15	상대오차 = -80.82	상대오차 = -81.06045	상대오차 = 26.59649
20	(4147325, 0.00497, 0.2297)	(4049601, 0.2512)	(5007992, 0.127)
30	상대오차 = -10.866	상대오차 = -12.96677	상대오차 = 7.630766
51	(4621533, 0.00583, 0.19418) 상대오차 = -0.675	(4606018, 0.2108) 상대오차 = -1.008375	(4740560, 0.1362) 상대오차 = 1.88317

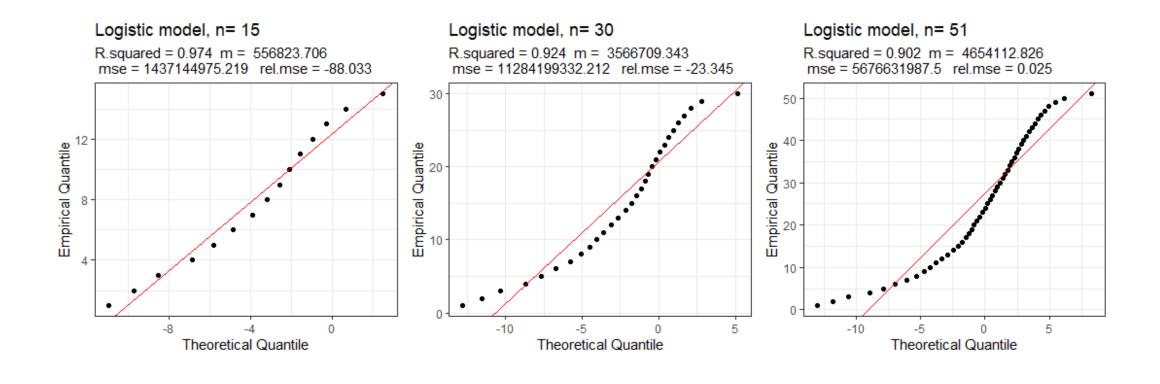
많은 자료를 이용할수록 총 수요량 값과 비슷하게 추정한 것을 확인할 수 있다.

#### 3. MSE, Q-Q plot 등을 이용하여 최적 예측모형 선택

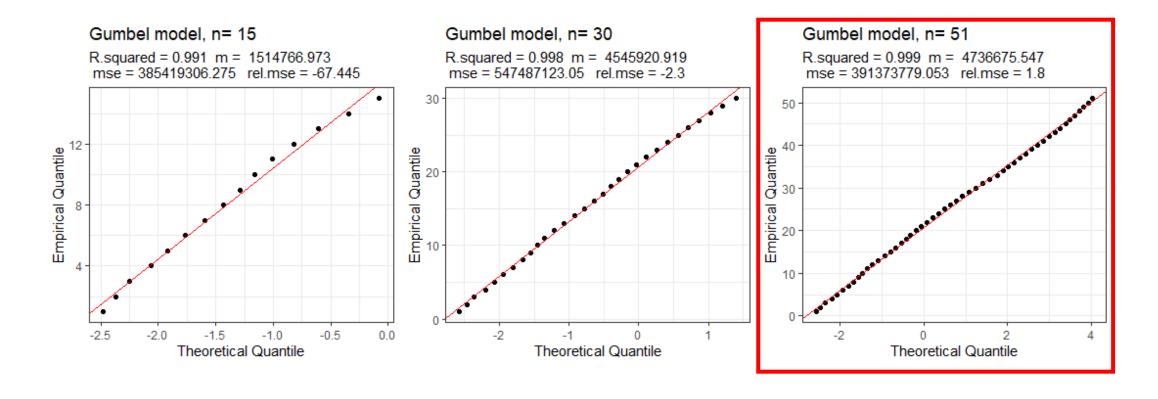




### 3. MSE, Q-Q plot 등을 이용하여 최적 예측모형 선택

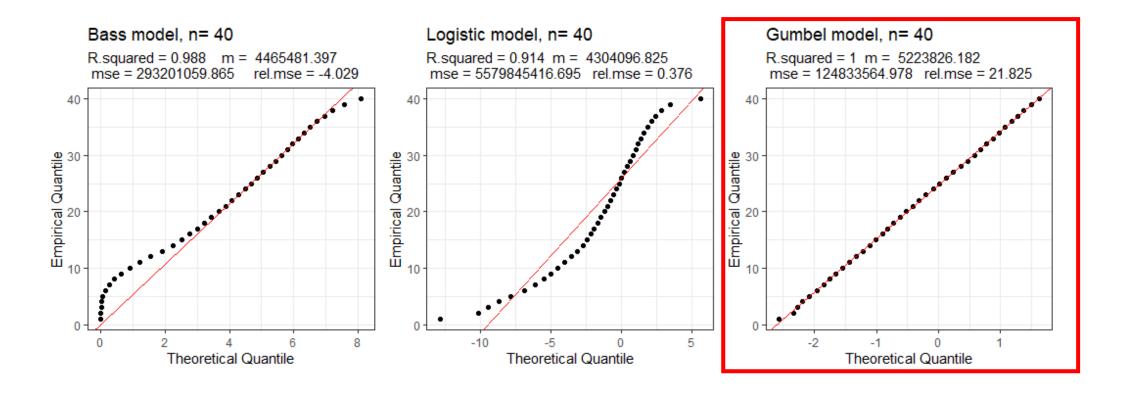


#### 3. MSE, Q-Q plot 등을 이용하여 최적 예측모형 선택

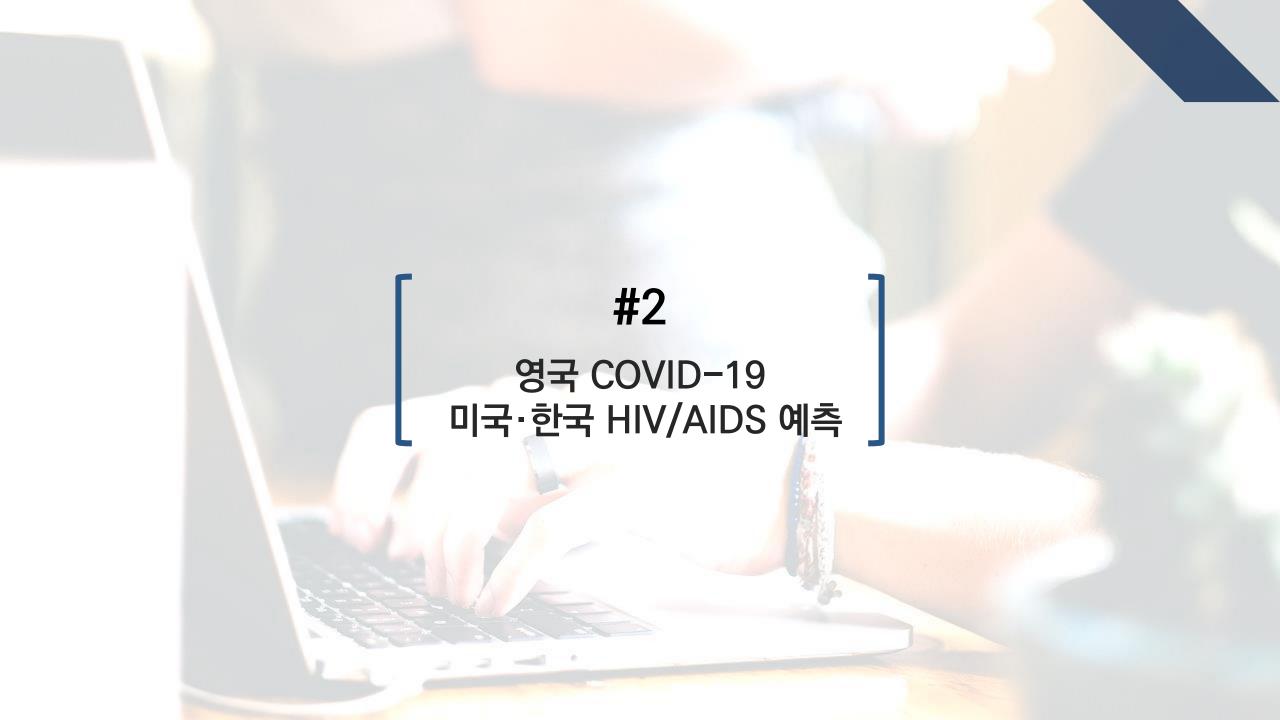


OLS 방법과 마찬가지로 n 이 클수록 상대오차가 작은 것을 확인할 수 있다. Q-Q plot을 살펴보면 Gumbel 분포를 가정하였을 경우 가장 직선 형태에 가깝다.

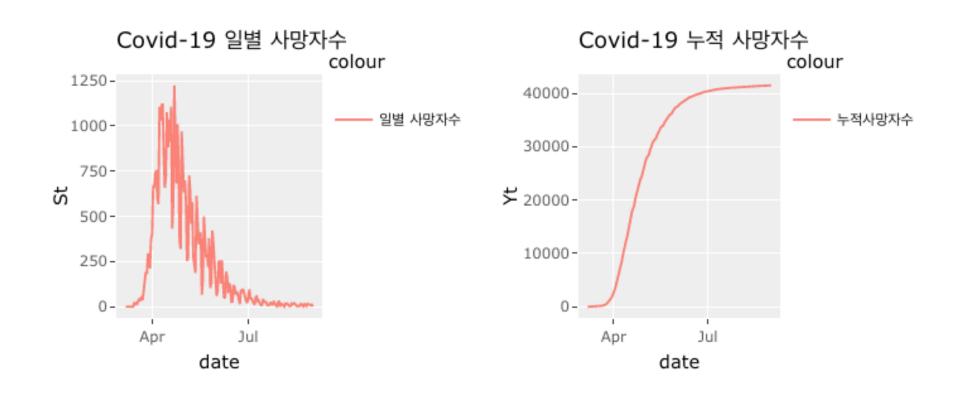
#### 4. 1M-DRAM 전체자료 (n=40)에 대해 예측한 m값 비교



Q-Q Plot을 살펴보면, Gumbel 모형에서 가장 직선을 잘 따르는 것을 확인할 수 있다.



## 1. 일별 Covid-19 사망자수 S(t), 누적사망자수 Y(t)에 대한 시계열 도표



5월의 사망자 수가 가장 컸고 그 이후로는 감소하는 경향을 보인다.

## 2. 모수 (p, q, m) 추정, m값의 상대 오차 계산 & 최적 모형 선택

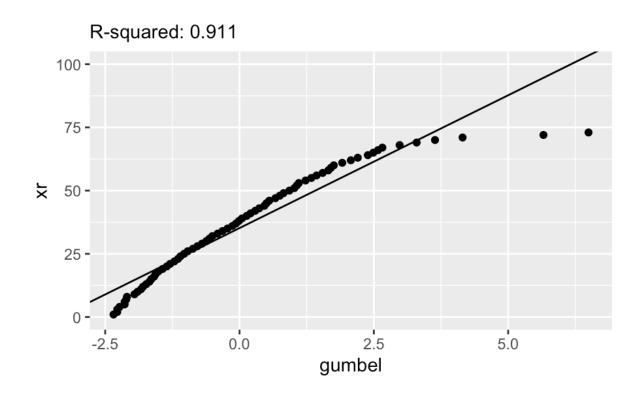
N	Bass (m,p,q)	상대오차
20	2.778e+32, 3.78e+28, 3.78e+28	inf
30	1.239e+04, 3.539e-04, 0.2726	-70
50	2.724e+04, 2.371e-03, 0.1466	-34

Ν	Logisitc(m,q) ; p=0	상대오차
20	-598.83, 0.204	-101.443
30	12065.71, 0.277	-70.925
50	26656.39, 0.158	-35.766

N	Gumbel(m,q); p=0	상대오차
20	4.0751, -0.0749	-99.99
30	135492.3, 0.0529	226.495
50	33872.14, 0.0776	-18.378

상대오차의 절대값이 가장 작은 Gumbel (N=50) Model이 가장 좋은 모델이다.

#### 3. m 추정값의 정확도 비교, 각 방법의 장단점 기술



Gumbel (N=50) 모델의 Q-Q Plot을 그렸을 때, M값이 너무 작아 gumbel : -log(-log(ur))로 하면, 114개의 데이터가 누락된다는 단점이 있다.

#### 4. 이탈리아 Covid-19의 최적 예측모형 찾기 & 정확도 비교

최초 n= 20, 30, 50일 자료 이용

N	Bass (m,p,q)	상대오차
20	12269.48, 0.005, 0.22	inf
30	19930.28, 0.005, 0.16	-65
50	28697.56, 0.0071, 0.092	-46

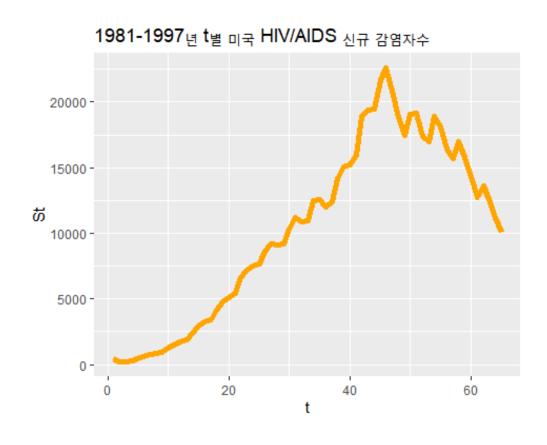
N	Logisitc (m,q) ; p=0	상대오차
20	10962.23, 0.257	-68.972
30	18815.09, 0.186	-46.745
50	27479.92, 0.118	-22.219

N	Gumbel (m,q) ; p=0	상대오차
20	23567.10, 0.086	-33.294
30	26598.35, 0.082	-24.714
50	30556.22, 0.069	-13.512

상대오차의 절대값이 가장 작은 Gumbel (N=50) Model이 가장 좋은 모델이다.

#### 미국 분기별 HIV/AIDS-감염자 자료 (1981-1997)

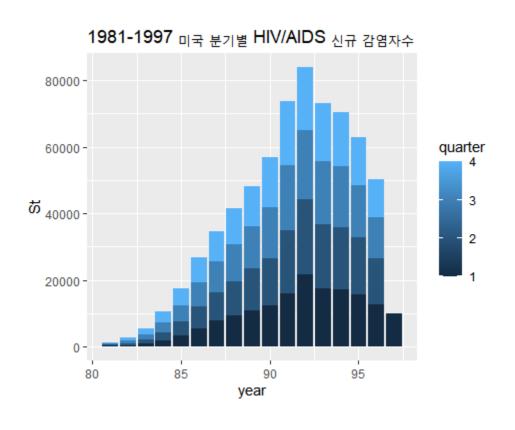
#### 1. 분기별 HIV/AIDS 감염자자료 S(t)에 대한 시계열 도표



t가 45-47일 때, 즉 1992년도 신규 에이즈 감염자수가 2만명을 넘었고, 1992년 2분기(t=46)에 최고점에 도달했다. 그 후부터는 감염자수가 현저하게 감소하는 추세임을 볼 수 있다.

#### 미국 분기별 HIV/AIDS-감염자 자료 (1981-1997)

#### 1. 분기별 HIV/AIDS 감염자자료 S(t)에 대한 시계열 도표



전반적으로 매년 각 분기별 신규 감염자수 비율은 균일하다. 즉 HIV/AIDS 감염자는 분기에 관계없이 비슷하게 비율로 발생한다.

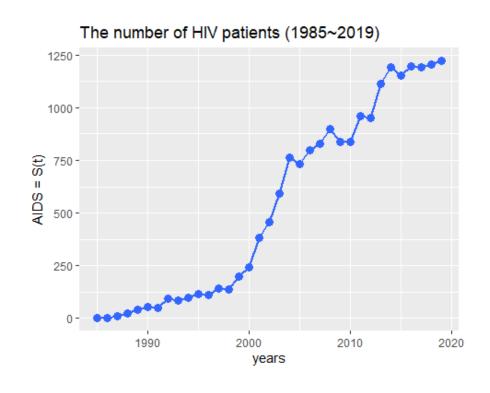
#### 미국 분기별 HIV/AIDS-감염자 자료 (1981-1997)

#### 2. 모수 (p, q, m) 추정, m값의 상대 오차 계산 & 최적 모형 선택

n	Bass	Logistic	Gumbel
n	$(\widehat{m}, \widehat{p}, \widehat{q})$	$(\widehat{m},\widehat{q})$	$(\widehat{m},\widehat{q})$
20	(116137, 0.00191, 0.214)	(88572, 0.2456)	(784235, 0.0503)
20	상대오차 = -90.923	상대오차 = -93.077	상대오차 = -38.705
40	(450452, 0.002, 0.12)	(403336, 0.1366)	(886444, 0.0461)
40	상대오차 = -64.793	상대오차 = -68.476	상대오차 = -30.716
45	(786740, 0.001525, 0.0965)	(773275, 0.1033)	(1000860, 0.0483)
65	상대오차 = -38.509	상대오차 = -39.562	상대오차 = -21.774

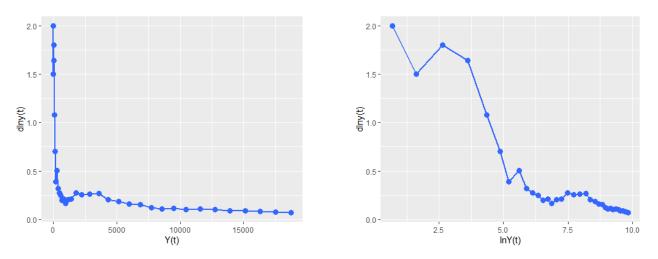
상대오차가 가장 작은 Gumbel 모형이 가장 타당한 모형이라고 할 수 있다. 또한 세 모형 다 n값이 커질수록 상대오차가 감소함을 볼 수 있다.

#### 1. 연도별 HIV/AIDS 감염자자료 S(t)에 대한 시계열 도표



국내 HIV 감염인의 숫자는 2000년까지는 완만하게 증가하다가 그 이후부터 가파르게 증가하는 추세를 보인다. 2015년부터는 증가추세가 약화된 것으로 보인다.

## 2. (dlnYt, Yt), (dlnYt, InYt) 산점도 & OLS (m,q) 추정

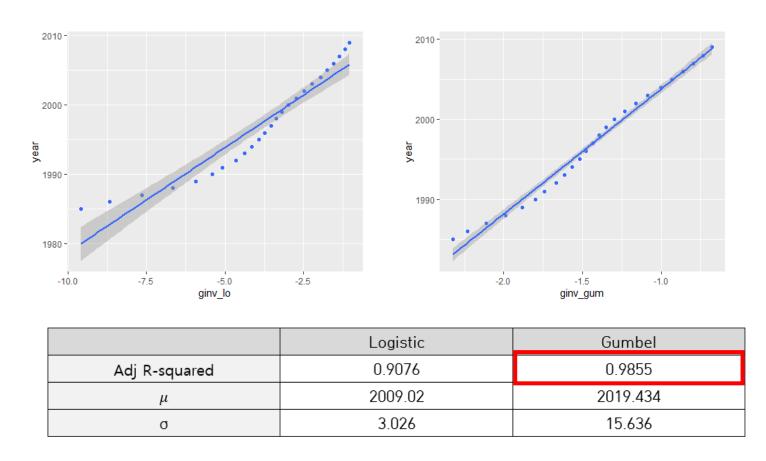


각 산점도에서 선형성을 찾기 힘들었기 때문에 dlny(t)와 y(t)를 직접적으로 이용하는 식이 아닌, s(t)와 y(t)를 이용하는 식을 사용해서 회귀 직선을 추정하였다.

	Regression	Adj R-squared	(q, m)
Logistic	$S(t) = 0.165y(t-1) - 5.6e-6 y(t-1)^2$	0.9855	(0.165, 29400)
Gumbel	S(t) = 0.66y(t-1) - 0.06y(t) Iny(t-1)	0.9928	(0.06, 54871)

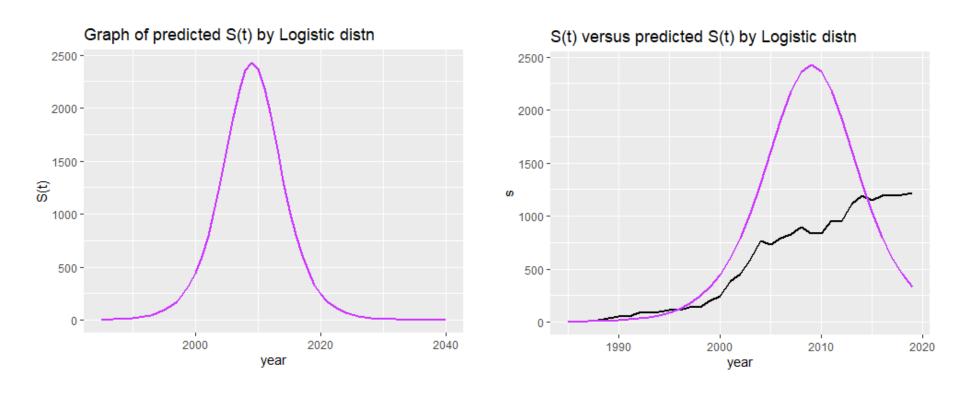
Adjusted R-squared 값이 더 높은 Gumbel 모델이 최적 모델이다.

#### 3. OLS m으로 그린 Logistic 및 Gumbel Q-Q plot & (μ,σ) 추정



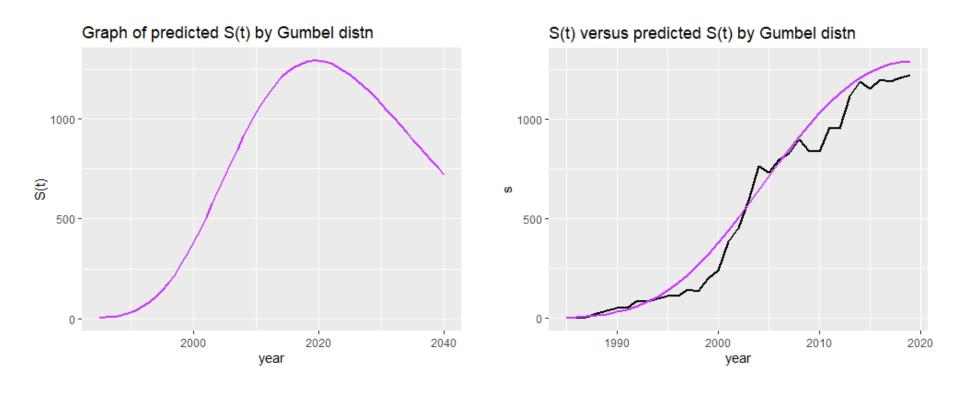
Q-Q Plot의 Adjusted R-squared를 비교했을 때, Gumbel분포의 값이 더 높다. 즉, Gumbel 분포가 1985년부터 2009년까지의 데이터를 더 잘 설명한다고 할 수 있다.

#### 4. S(t)의 예측값을 추정(1985~2040년) & 의미 설명

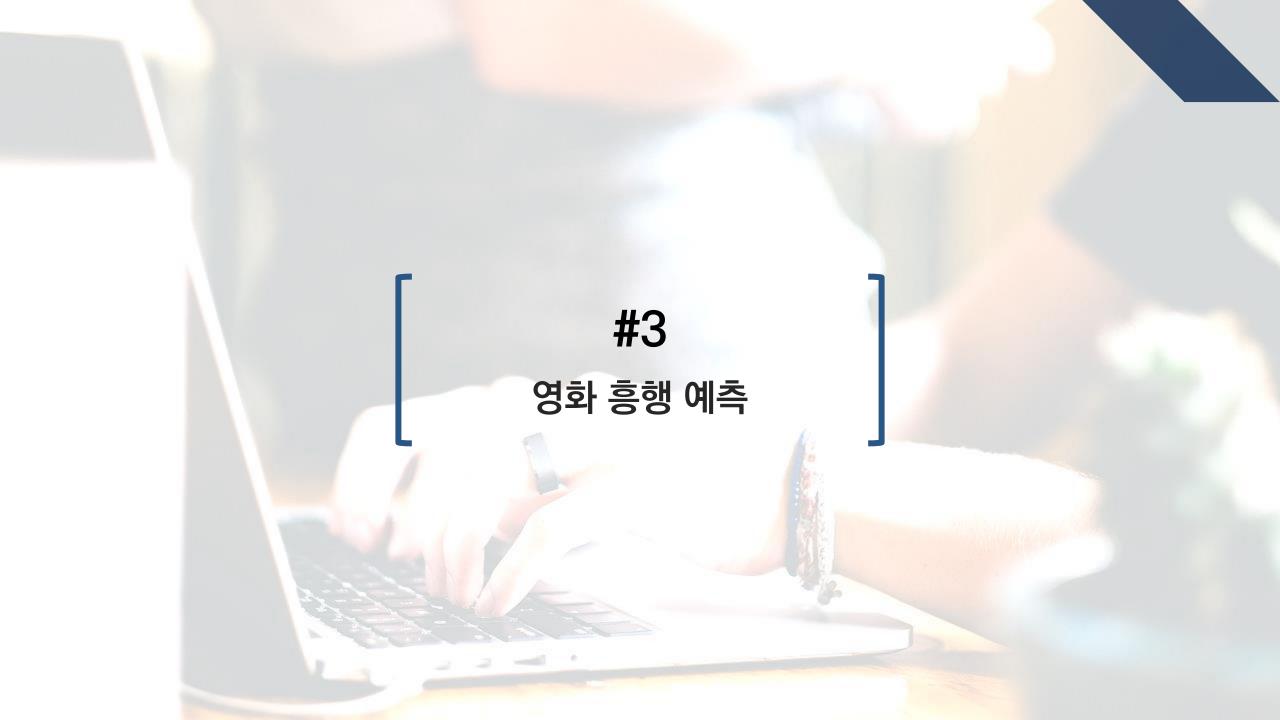


Logistic 분포로 구한 예측 값은 약 2010년에 최댓값을 기록하고 점점 감소함을 볼 수 있다. 이를 실제 관측 값과 비교해보면, logistic 분포는 S(t) 값을 예측하기에 적절하지 않음을 알 수 있다.

#### 4. S(t)의 예측값을 추정(1985~2040년) & 의미 설명

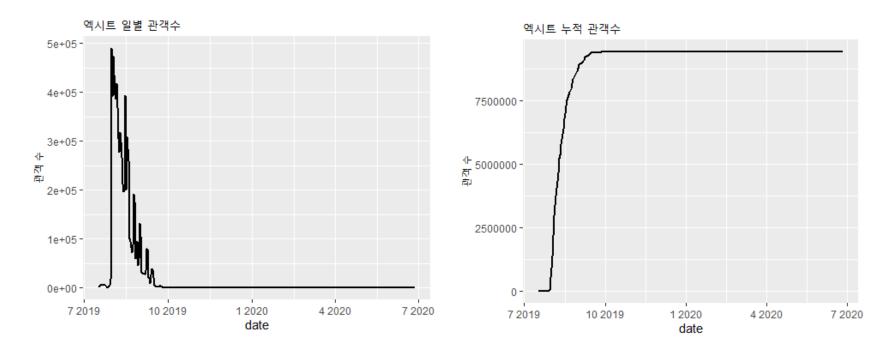


Gumbel 분포로 구한 예측 값은 약 2020년 최댓값을 기록하고 그 이후 점점 감소하는 경향을 보인다. 실제 값과 일치하는 경향을 보이는 Gumbel 분포가 적절한 분포임을 확인할 수 있다.



## 1. 일별 관객수S(t) 및 누적 관객수Y(t) 시계열 도표

#### # 엑시트

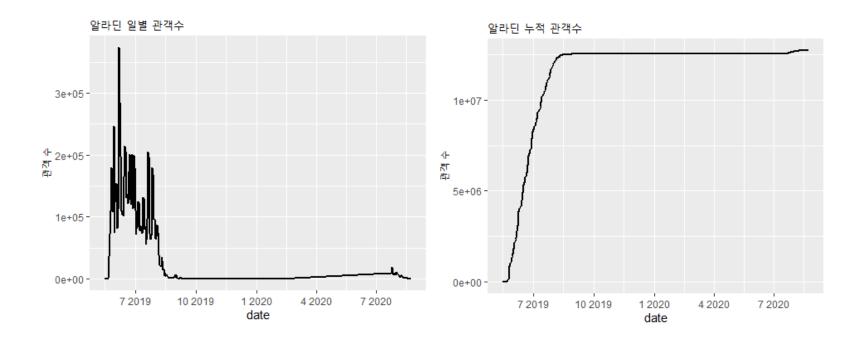


약 2019년 9월을 기점으로 증가하지 않았다.

#### 영화 흥행 예측

## 1. 일별 관객수S(t) 및 누적 관객수Y(t) 시계열 도표

#### # 알라딘



약 2019년 8월을 기점으로 증가하지 않았다.

### 2. m값의 상대 오차 계산 & 최적 모형 선택

개봉 후 1주, 2주 및 4주 간의 흥행 자료를 이용

#### # 엑시트

총 관객수 추정 (실제 m : 9426161)					
bass logisic gumbel Exponent					
Week1	15303	15071	15264	21640	
Week2 2355667 2297368 2648178 -275878					
Week4	6794898	6640789	7073915	-5423895	

상대오차					
bass logisic gumbel Exponent					
Week1	-99.84	-99.84	-99.84	-99.77	
Week2	-75.01	-75.63	-71.91	-102.93	
Week4	-27.91	-29.55	-24.95	-157.54	

Gumbel 모형의 성능이 가장 뛰어나고 Exponential 모형으로 추정한 m의 상대오차의 절대값이 가장 크다.

### 2. m값의 상대 오차 계산 & 최적 모형 선택

개봉 후 1주, 2주 및 4주 간의 흥행 자료를 이용

#### # 알라딘

총 관객수 추정 (실제 m : 12723775)					
bass logisic gumbel Exponent					
Week1	562974	-20401			
Week2 1528417 1487495 1640016 -675373					
Week4	8088579	5704853	7697641	-1775874	

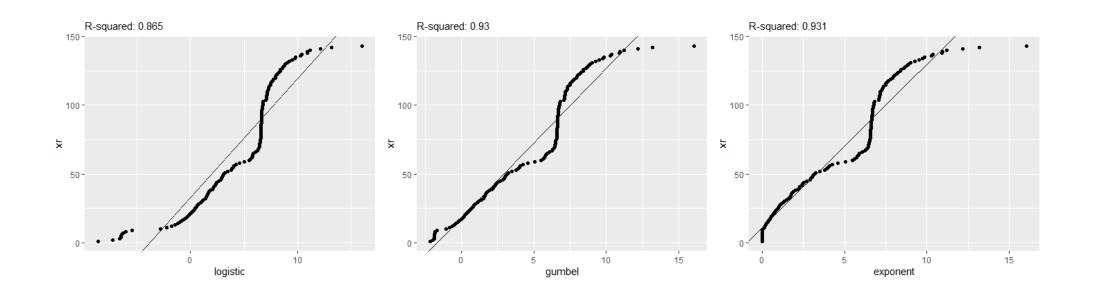
상대오차				
	bass	logisic	gumbel	Exponent
Week1	-97.08	-97.31	-95.58	-100.16
Week2	-87.99	-88.31	-87.11	-105.31
Week4	-36.43	-55.16	-39.50	-113.96

Bass 모형의 성능이 가장 뛰어나고 Exponential 모형으로 추정한 m의 상대오차의 절대값이 가장 크다.

## 3. 실제 총 관객수 m을 이용하여 Q-Q plot 작성

개봉 후 1주, 2주 및 4주 간의 흥행 자료를 이용

#### # 엑시트



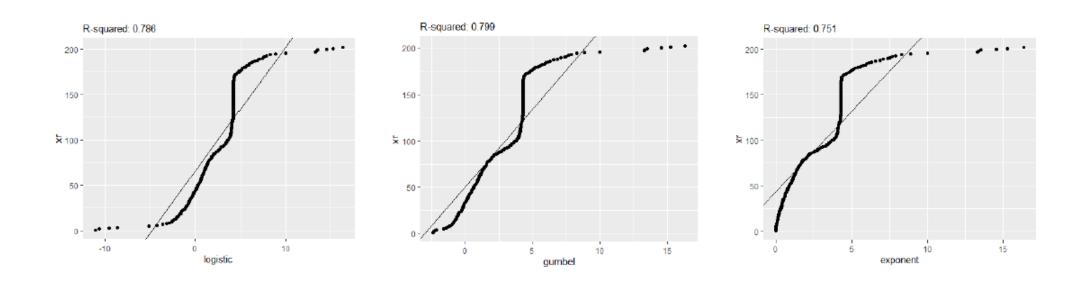
Q-Q plot을 그려본 결과 엑시트의 경우 Gumbel과 Exponential 모형이 가장 적절하다.

#### 영화 흥행 예측

## 3. 실제 총 관객수 m을 이용하여 Q-Q plot 작성

개봉 후 1주, 2주 및 4주 간의 흥행 자료를 이용

#### # 알라딘



Q-Q plot을 그려본 결과 알라딘의 경우 Gumbel 모형이 가장 적절하다.



