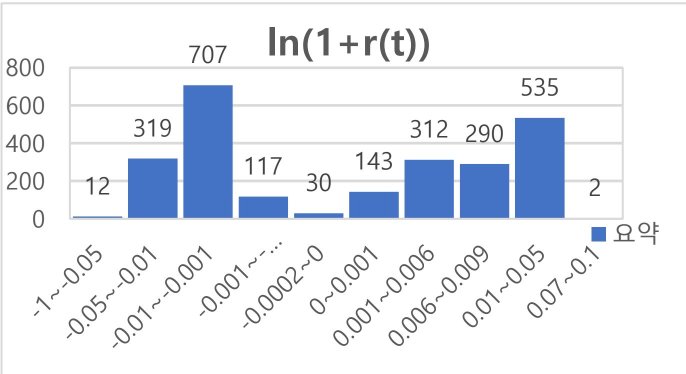
이통과제#3

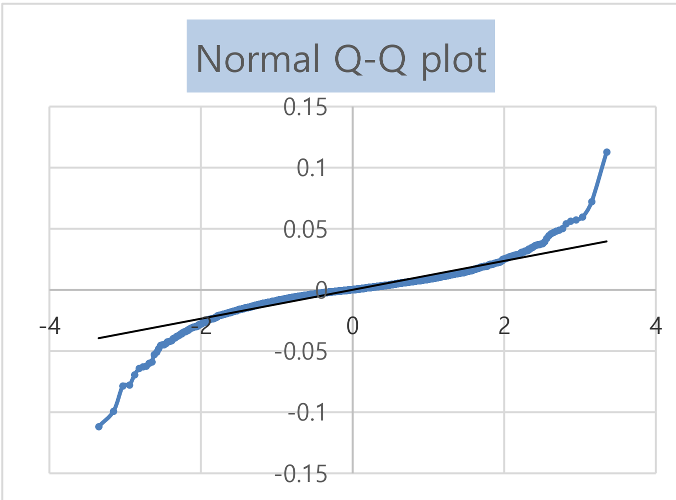
182STG12 오혜윤

***자세한 그래프나 자료는 엑셀에 있습니다.**

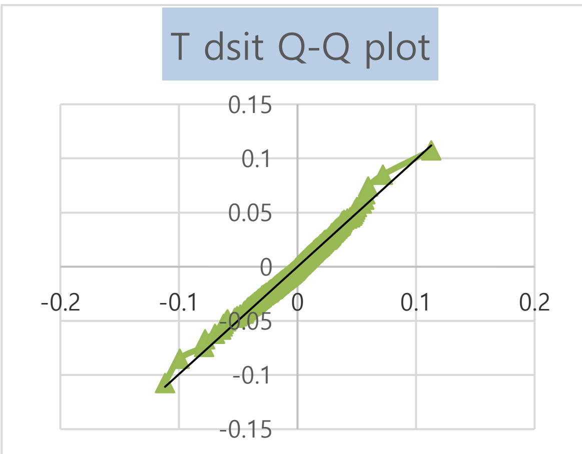
Part 1) KOSPI 일별수익률의 VAR(Value at Risk) 값 계산

a) qqplot을 그려보았을 때, -2~2 범위에서 추세선에 가깝게 그려지나, 그 밖의 범위에서는 많이 벗어나는 것으로 보아, 정규성이 잘 성립하지 않는다고 보여집니다.





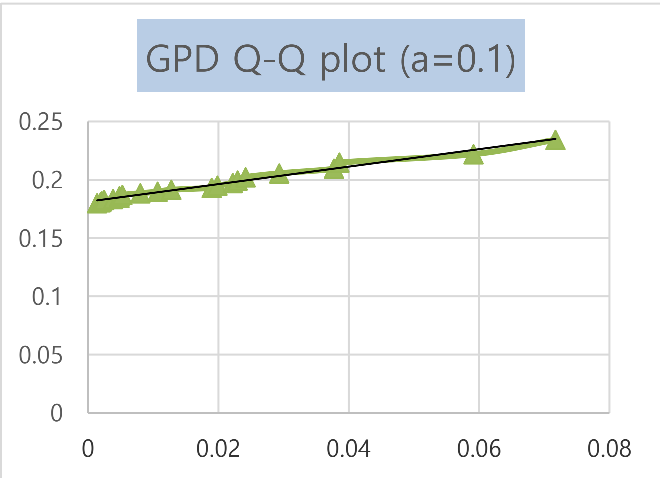
b) logistic, laplace , t distribution qqplot을 그려본 결과, 자유도=3일 때의 t 분포의 qqplot이 가장 직선 (추세선)에 가까운 그래프를 보이는 것을 알 수 있었습니다. 따라서 rt의 분포를 자유도 3일 때의 t분포로 두었습니다. (다른 qqplot 그래프는 엑셀에 있습니다.)



이에 따라, 자유도=3인 t분포를 rt의 분포로 보고, var(alpha)=varx값을 구했습니다. 그 결과, 아래와 같은 표를 얻을 수 있었습니다.

alpha	varx
0.01	0.0378
0.004	0.058
0.001	0.0785
0.0004	0.1

c) a에 따라, qqplot을 그려본 결과, 추세선에 가장 가깝게 그려지는 것은 a=0.1일 때임을 알 수 있었습니다.



따라서, a=0.1일 경우, GPD의 모수는 다음과 같습니다.

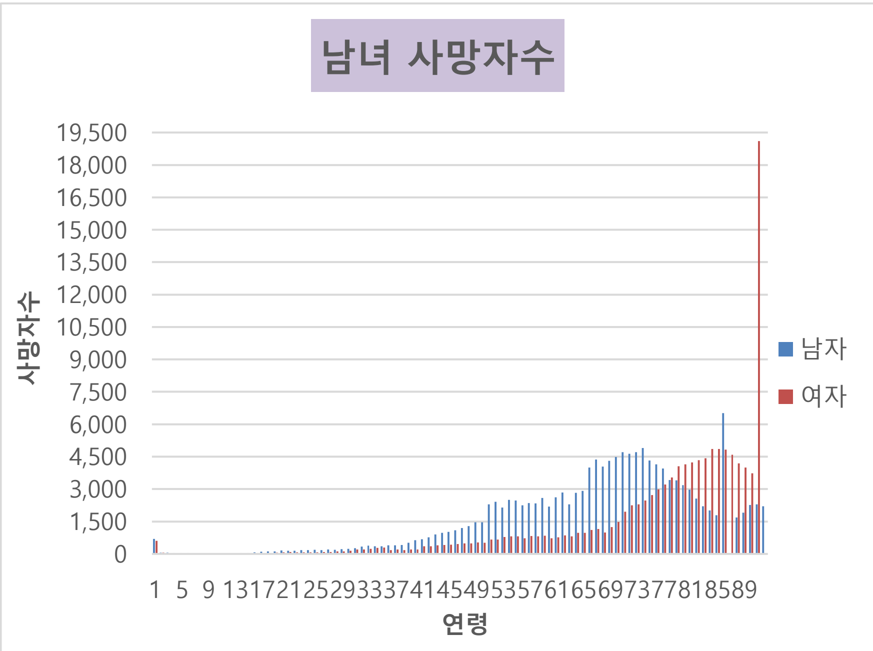
a	r	sigma
0.1	0.17902	0.0179017

d) alpha에 따른 var(alpha)를 구한 결과는 다음과 같습니다.

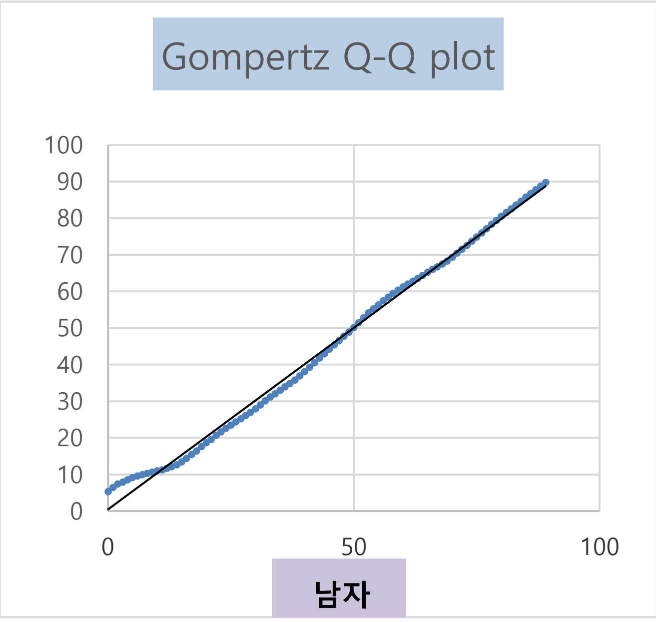
alpha	ALPHA0	ALPHA1	VAR_ALPHA
0.01	0.00848	1.1790476	0.037094505
0.004		0.471619	0.054401784
0.001		0.1179048	0.086626324
0.0004		0.0471619	0.112762743

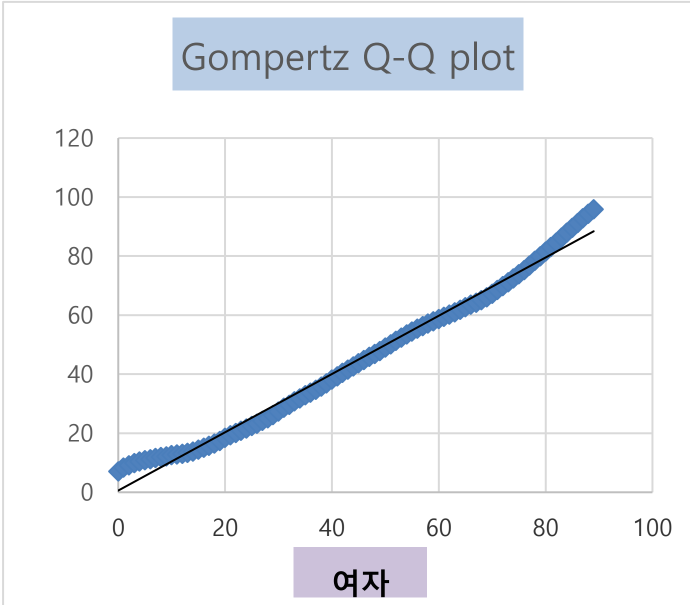
Part 2) 인간의 수명분포

a) 남녀 사망자수를 겹쳐 그려 보았을 때, 여자의 평균 사망 연령이 남자에 비해 더 높은 것을 알 수 있었습니다. 이것은 남자보다 여자가 더 오래사는 것을 의미합니다. 특히, 여자의 경우, 90세 이상의 사망 비율이 월등히 높았습니다. 연령이 낮을 때, 남자의 경우 여자보다 사망자수가 많았습니다.



b) 남자의 Gompertz qqplot을 그려본 결과, 끝점을 제외하고 대부분 직선에 가까운 것을 알 수 있었습니다. 반면에, 여자의 Gompertz qqplot을 그려본 결과, 남자 qqplot에 비해 끝점이 잘 맞지 않고, 직선에 가깝지 않은 것을 볼 수 있었습니다.



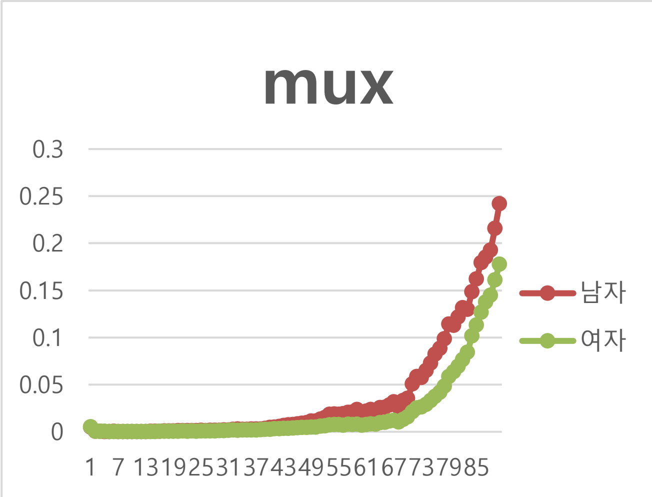


이에 따라, 모수를 추정한 결과표는 다음과 같습니다.

남자	
MU	75.01863
SIGMA	13.03486

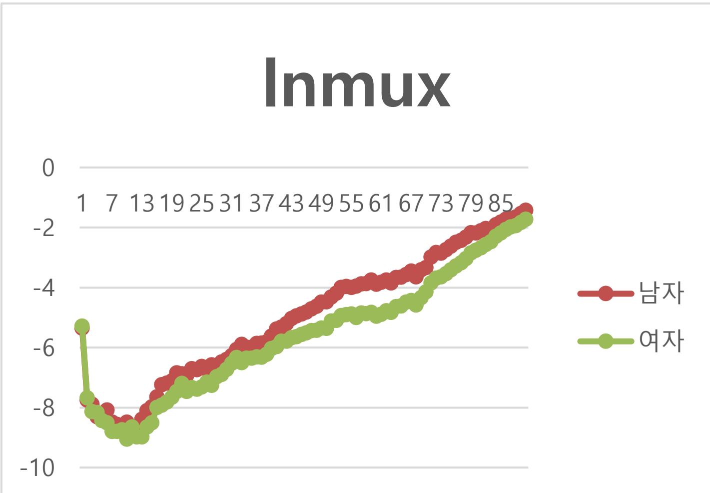
여자	
MU	86.75041
SIGMA	15.08155

c) mux의 그래프를 볼 때, 연령 별 사망률은 남자의 경우, 여자의 경우보다 좀 더 급격하게 증가함을 알 수 있었습니다. 그리고 여성에 비해 사망률이 증가하는 시기가 훨씬 빠른 것을 알 수 있었습니다. 또한, 사망률의 절대적인 크기로 보았을 때, 남자의 경우가 여자의 경우보다 훨씬 사망률이 큰 것을 알 수 있었습니다. (그래프는 아래에 있습니다.)



lnmux의 그래프를 볼 때, mux 그래프의 경우와 마찬가지로, 남자의 경우 사망률이 좀 더 빨리 증가함을 알 수 있었습니다. 또한, 기울기로 보아, 남자의 경우 사망률이 좀 더 급격하게 증가함을 알 수 있었습니다.

mux그래프와 lnmux그래프를 그려 보았을 때, mux 그래프보다 기울기를 좀 더 자세하게 볼 수 있어서 lnmux의 그래프가 편리하고, 상대적으로 비교하기 쉽다는 것이 장점이었습니다.



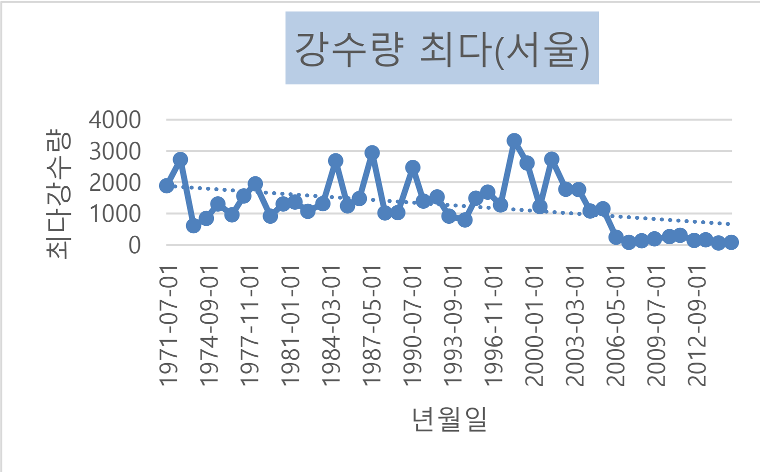
사망률 그래프는 지수 분포를 따르고, gompertz 법칙이 성립함을 볼 수 있었습니다. gompertz 법칙에 따라 추정한 모수는 다음과 같습니다.

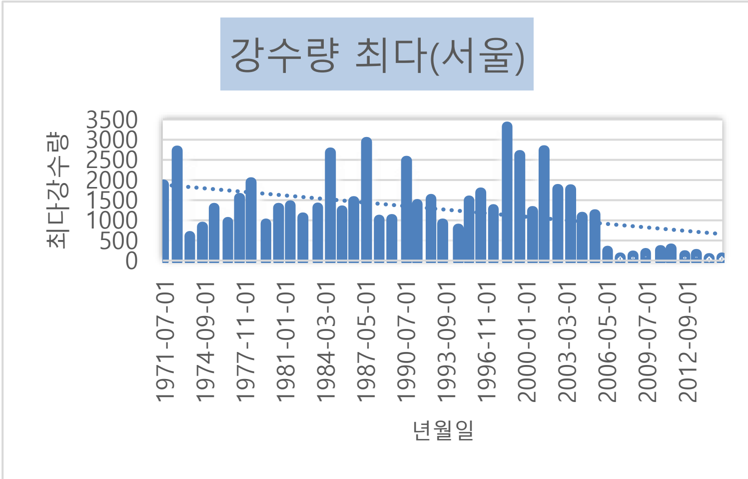
남자	
a	-8.5623
b	0.080194

여자	
a	-9.14767
b	0.078577

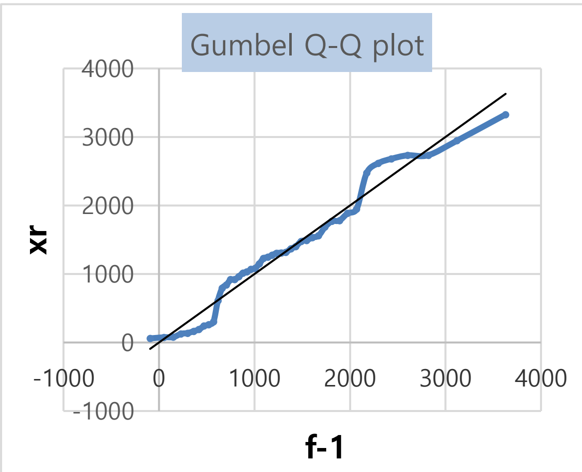
Part 3) 기상변수 극값(최대값) 분포 및 홍수피해액 분포

a) 연중 최대일 강수량의 시계열 도표와 히스토그램은 다음과 같습니다.





b) gumbel qqplot과 모수를 추정한 결과는 다음과 같습니다.

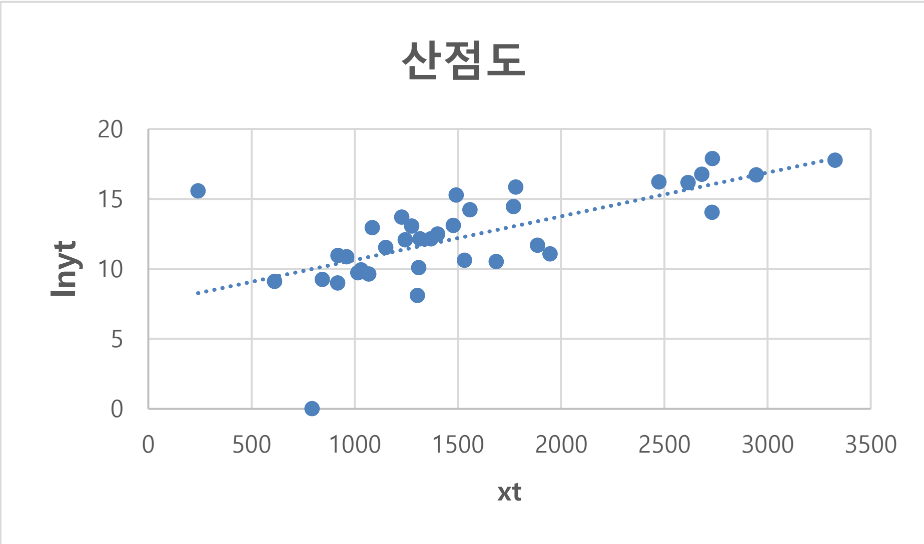


mu	875.5276
sigma	721.3456

c) T년 최대 강수량 UT는 다음과 같습니다.

UT(T=50)	3690.174
UT(T=100)	4193.825
UT(T=200)	4695.638

d) lnyt와 xt의 산점도는 아래와 같습니다.

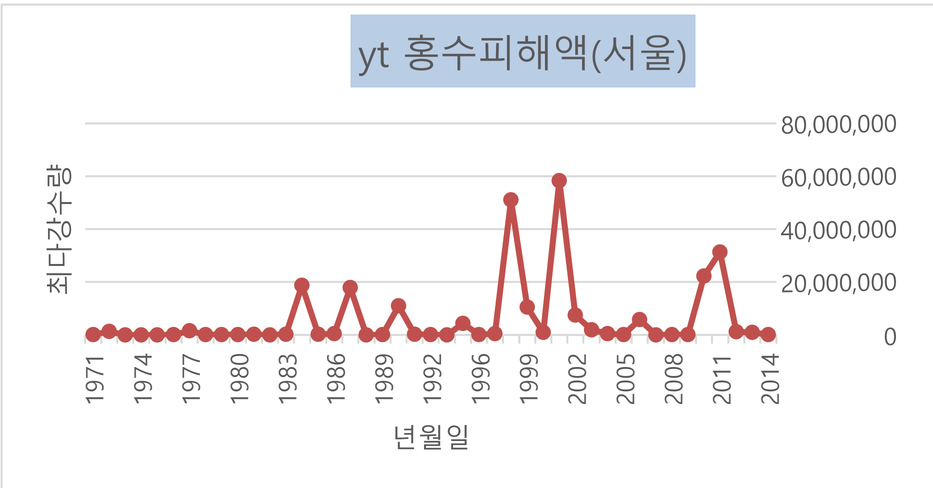


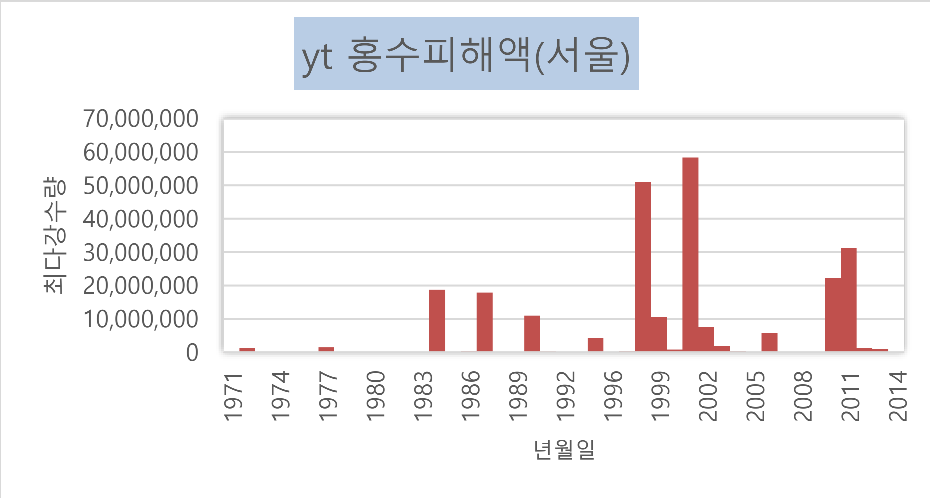
회귀모형에 따른 모수 추정 결과는 아래와 같습니다.

a	7.507748
b	0.003123

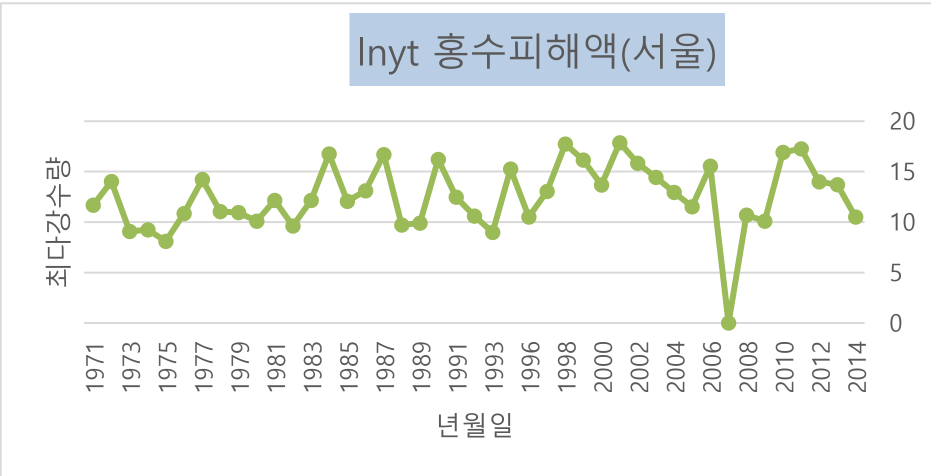
홍수피해액과 강수량 최다는 양의 상관관계를 가짐을 알 수 있습니다. 산점도와 회귀 모형은 강수량이 증가함에 따라 홍수 피해액도 증가함을 나타냅니다. 최대강수량이 1씩 증가 할 때마다, 홍수 피해액은 0.003배씩 증가합니다.

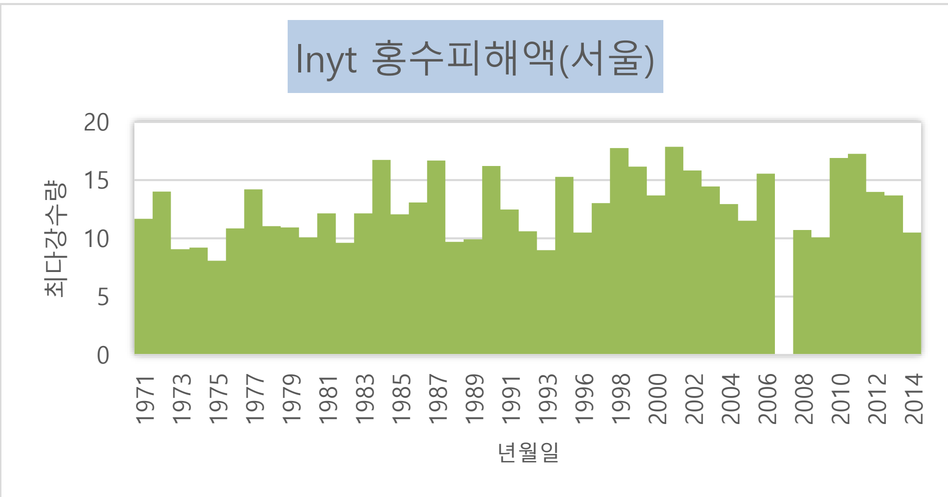
e) yt의 시계열 도표와 히스토그램은 아래와 같습니다.



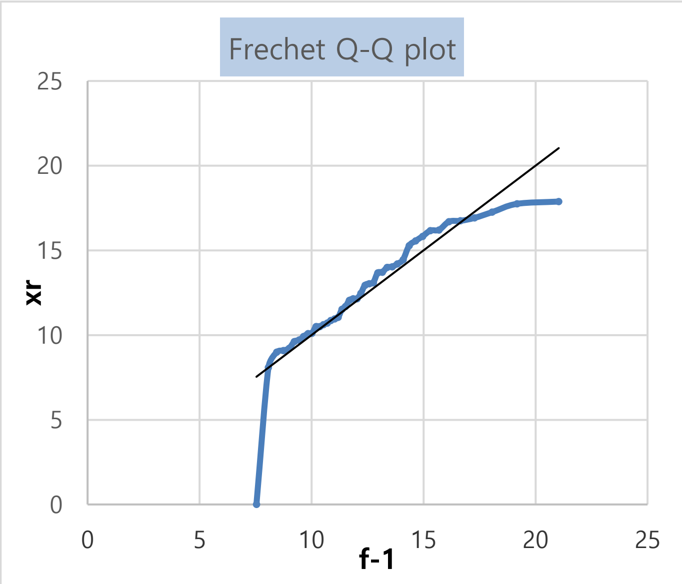


lnyt의 시계열 도표와 히스토그램은 아래와 같습니다.





frechet qqplot의 결과는 다음과 같습니다.



이에 따라, 모수를 추정한 결과표는 다음과 같습니다.

mu	11.05867
sigma	2.644772

Part 4) [EVT에 근거한 추론의 타당성 검토]

a) 조선시대 자료의 경우, 1777년부터 1907년까지의 데이터 중 1907년 자료가 없기 때문에, 1907년 자료를 제외하고 분포를 구했습니다. (1906년까지의 데이터만 사용했습니다.)

log normal, log logistic, frechet, gumbel 등의 분포의 qqplot을 그려본 결과, log normal 일 때의 qqplot 상에서 가장 직선에 가까움을 확인했습니다. 따라서, log normal 분포를 이용하여, 최대 강수량을 구했습니다. 최대강수량 표는 다음과 같습니다.

UT(T=50)	2.759443
UT(T=100)	2.8327
UT(T=200)	2.899743

b) 앞서 구했던 최대 강수량은 강수량의 크기를 의미했고, 이번에 구한 강수량은 cm를 의미하기 때문에 단위를 먼저 맞춰주어야 했습니다. 서울의 면적은 605.21km^2이다. 강수량 1㎜는 1㎡의 면적에 1ℓ의 양의 강수가 내린 것이 기 때문에 단위를 맞춰주면, cm*0.01*605.21*1,000,000이 강수량이 됩니다. 하지만, 이는 앞서 구했던 최대 강수량보다 월등히 많은 양입니다. 결과적으로는 T가 증가할수록 모두 최대 강수량이 증가함을 보였습니다. 차이점은 앞서 구한 최대 강수량이 정확히 어떤 면적을 통해 구한 것이기 모르기 때문에, 비교할 수 없었고, 조선시대 강수량 자료가 월등히 많은 강수량을 보였다는 것을 알 수 있었습니다.

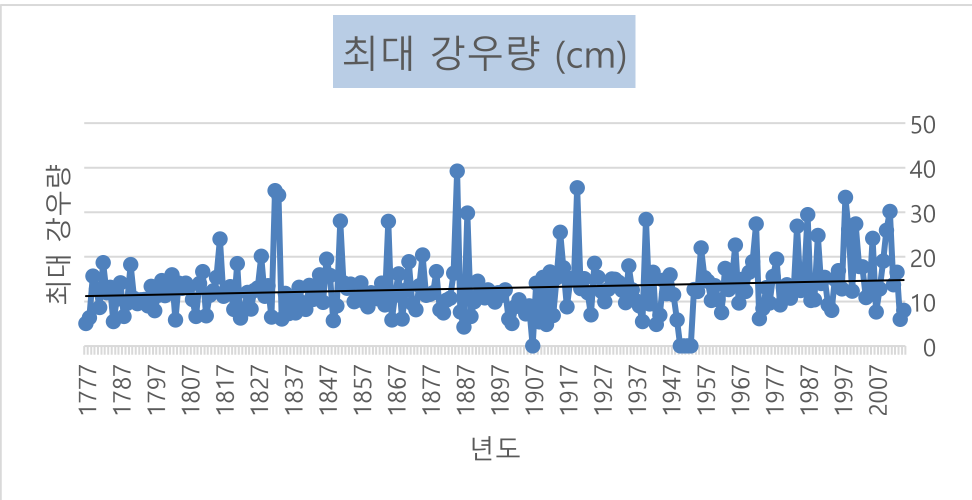
UT(T=50)	3690.174
UT(T=100)	4193.825
UT(T=200)	4695.638

<앞서 구했던 최대강수량>

UT(T=50)	16625922
UT(T=100)	17067299
UT(T=200)	17471245

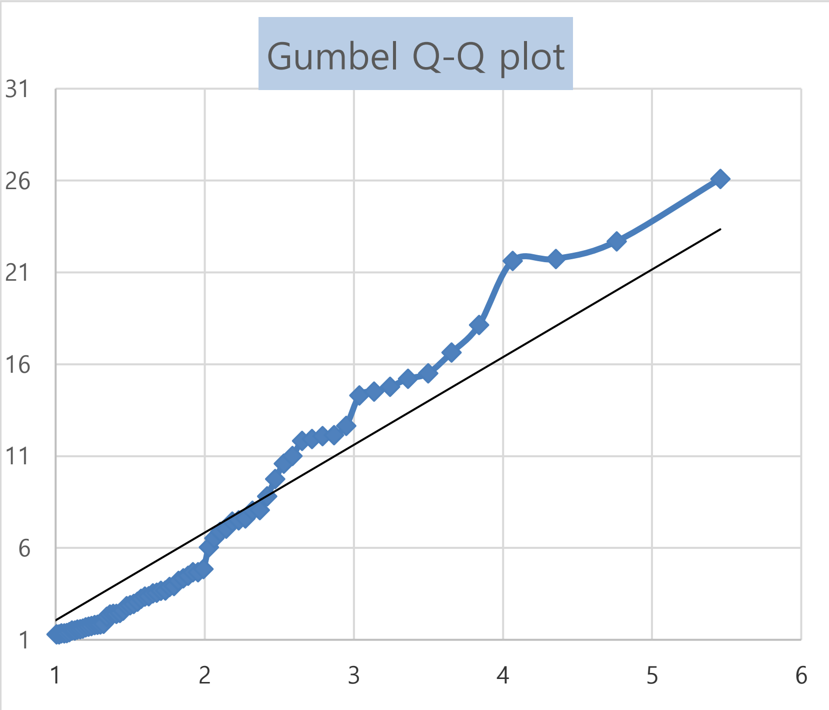
<이번에 구한 강수량>

c) 1777년부터 2015년의 자료의 경우, 일부 소실된 데이터가 있어, 그 데이터들을 제외하고, 모수를 추정했습니다. 따라서, 전체 자료의 시계열 도표와 모수를 추정한 결과는 다음과 같습니다.



alpha	11.13762
beta	0.018023

d) 잔차의 qqplot 과 모수 추정은 다음과 같습니다.



잔차	
mu	– 2.71388
sigma	4.775207

e) mu2016값과 일최대강수량이 30cm이상 될 확률은 다음과 같습니다.

mu2016	12.8395
prob>30	0.02712

f) 모수와 확률 추정 표는 다음과 같습니다.

MLE		
alpha	11.13762	
sigma	4.775207	
beta	0.018023	
mu2016	15.5533	
prob>30	0.04738	

e)와 같은 방법으로 추정했을 경우, 좀 더 쉽게 식을 써서 구할 수 있었지만, mle의 solver의 경우는 생소해 쓰는 것이 편리하지는 않았습니다. 하지만 mle의 경우에 선형성을 만족시키지 않아도 구할 수 있어서 더 정확한 값을 구할 수 있는 것이 장점이라 생각합니다.