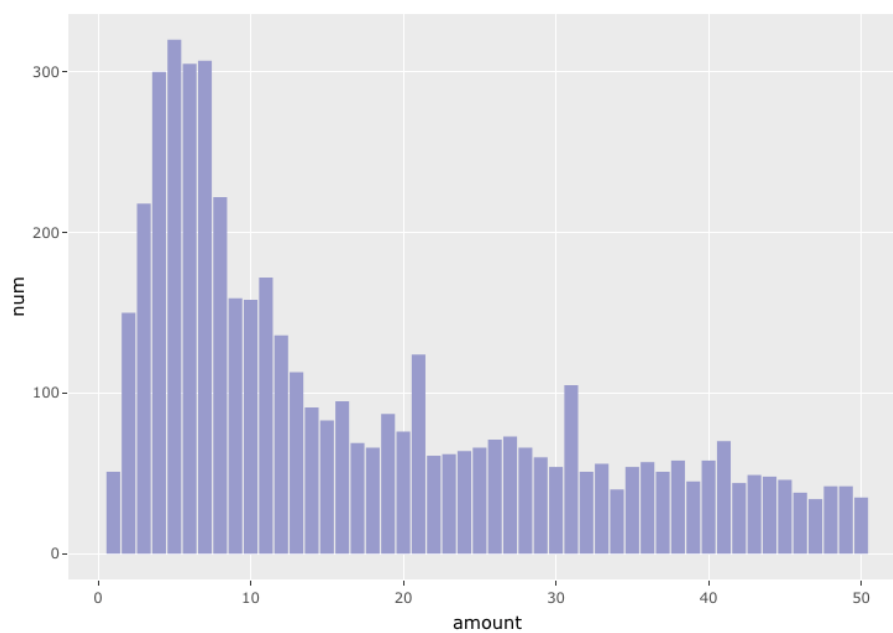


## 이론통계학2 – project #2 자기부담금 및 보상한도에 따른 상해/화재 보험료계산

2조 : 강아미 고유정 윤보인 이혜린 홍지원

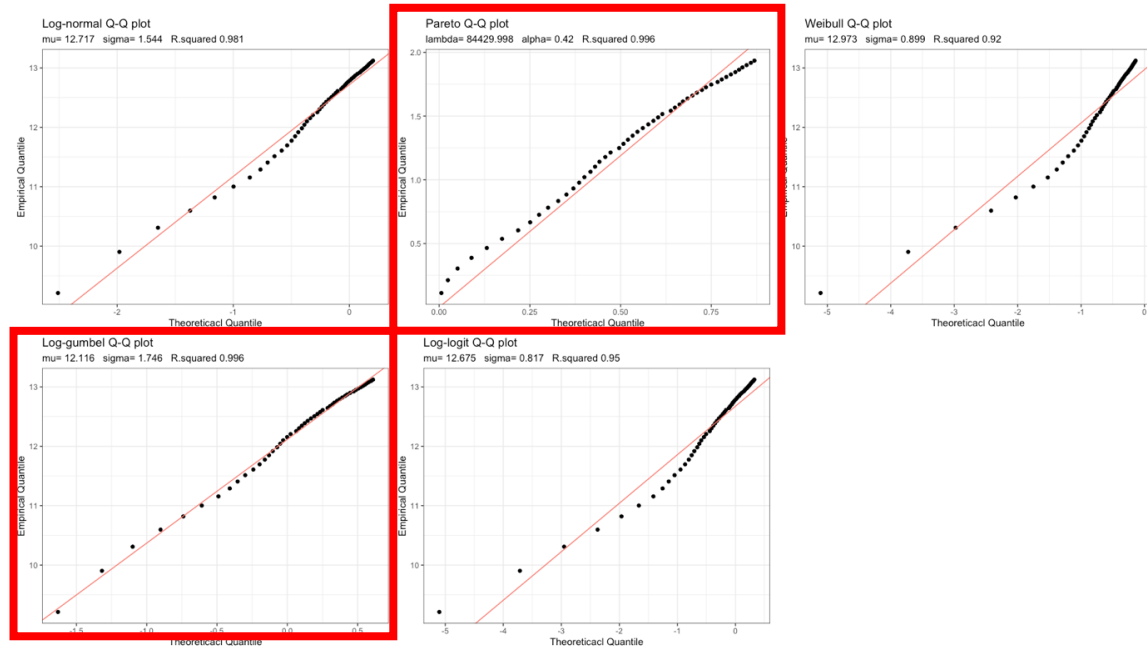
### <Part1> 상해보험

1. 아래 표 2 및 Excel file에 수록된 만원 단위로 계급화된 자동차 사고 치료비(x)의 histogram을 그려보시오. ( $X < 50$ )



➔ 뒷장 계속

2. 위 histogram 및 Q-Q plot을 이용하여 사고 손실액(X)에 가장 적합한 확률분포 2개를 찾아보시오. (Hint : 보상한도 50만원에서 censored자료에 대해 아래 5개의 각 모형들에 대해 Q-Q plot을 그리고 가장 적합한 모형 2개를 선택한 후 회귀분석으로 해당 모수를 각각 추정함)



➔ 적절한 분포는  $R^2$  값이 0.99와 0.996으로 가장 높은 값을 가지는 Pareto분포와 Log-Gumbel 분포가 됩니다. (값을 10000원 단위로 넣었다.)

	모수	$R^2$
Log-normal	(mu, sigma) = (12.717, 1.544)	0.98
Pareto	(lamda, alpha) = (84429.998, 0.42)	0.996
Weibull	(mu, sigma) = (12.973, 0.899)	0.92
Inverse Weibull;	(mu, sigma) = (12.116, 1.746)	0.996
Loglogistic	(mu, sigma) = (12.675, 0.817)	0.95

3. 위에서 선택된 2개의 최적모형을 이용하여 자기부담금(Deductible)을 A=0,10,20만원 보상한도 (Limit)를 B=50, 100, 200, 500, 1000만원으로 각각 조정하였을 때 적정 보험료를 각각 계산하십시오.

(Hint:  $E(N)$  = 사고건수 / 계약건수,  $Y = \min[(X - A)^+, B]$ , 보험료 =  $E(N) * E(Y)$  )

\*Pareto

	50만원	100만원	200만원	500만원	1000만원
A=0	9381	15378	24548	44233	67999
A=10만원	8119	13826	22752	42190	65819
A=20만원	7338	12799	21502	40703	64198

\*Log-Gumbel

	50만원	100만원	200만원	500만원	1000만원
A=0	9573	15231	23391	39506	57257
A=10만원	8113	13443	21336	37193	54810
A=20만원	7199	12251	19901	35513	52999

➔ 보상한도(A=0, 10만원, 20만원)를 늘릴수록 적정보험료는 감소하는 패턴을 확인할 수 있다.

4. 공제가 없고(A=0) 보상한도 (B=50) 가 50만원일 경우 위에서 계산한 두 가지 새로운 보험료를 적용했을 때 손해율((총보험금/총보험료)\*100)을 각각 계산하여 기존의 보험료의 손해율과 서로 비교해 보시오.

	기존 손해율 보험료=3130	Pareto 보험료=9380.94	Log Gumbel 보험료=9572.675
손해율	305.821	102.039	0.99

➔ Pareto Model과 Log Gumbel모델에 각각 새로운 보험료를 적용했을 때, Pareto Model의 손해율은 102%, Log-Gumbel의 손해율은 0.99입니다.

5. Excel-VBA / R-Shiny 등 도구를 이용하여 임의의 공제액(A), 보상한도(B)에 따른 적정보험료를 계산하는 Application을 만드시오.

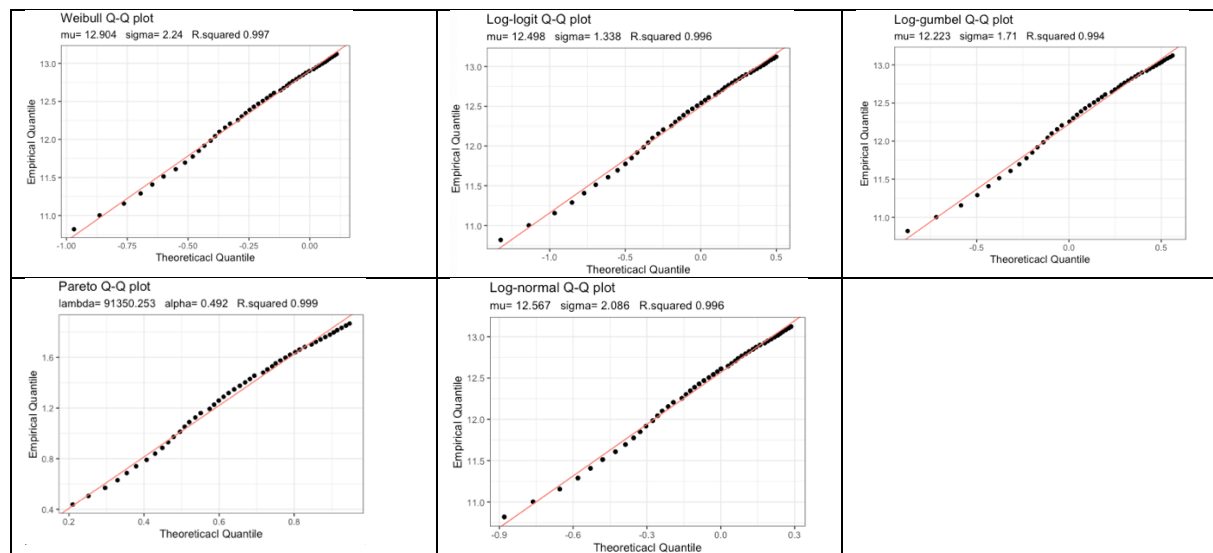
→ Markdown 참조

6. 만약 자료에서 5만원의 자기 부담금이 있어 5만원의 자기 부담금이 있어 5만원 이하의 사고 자료가 없는 경우 아래의 절단된(truncated)자료를 가지고 손해액(X)의 분포를 추정하고 5만원 이하의 사고건수(n1) 및 총 사고 건수(m=7404 + n1)를 추정하고 이를 표 2)의 실제값 n1, m 및 자기부담금이 없는 경우 2)에서 추정한 모수추정값과 서로 비교해보시오.

\*N1=1039

[ 각 분포별  $R^2$ 를 최대화하는 n1을 구하기 위한 모수 추정 ]

분포	N1	모수	$R^2$
Log-normal	1731	(mu, sigma) = (12.567,2.086)	0.996
Pareto	1091	(lambda, alpha) = (91350.253,0.492)	0.999
Weibull	30418	(mu, sigma) = (12.973,0.899)	0.92
Inverse Weibull	742	(mu, sigma) = (12.223,1.71)	0.994
Loglogistic	1569	(mu, sigma) = (12.498,1.338)	0.996



→ weibull은 5만원 이하의 truncated data를 잘 추정하지 못한다.

[ 2)에서 추정한 모수 추정값 ]

	모수	$R^2$
Log-normal	$(\mu, \sigma) = (12.717, 1.544)$	0.98
Pareto	$(\lambda, \alpha) = (84429.998, 0.42)$	0.996
Weibull	$(\mu, \sigma) = (12.973, 0.899)$	0.92
Inverse Weibull; Frechet;	$(\mu, \sigma) = (12.116, 1.746)$	0.996
Loglogistic	$(\mu, \sigma) = (12.675, 0.817)$	0.95

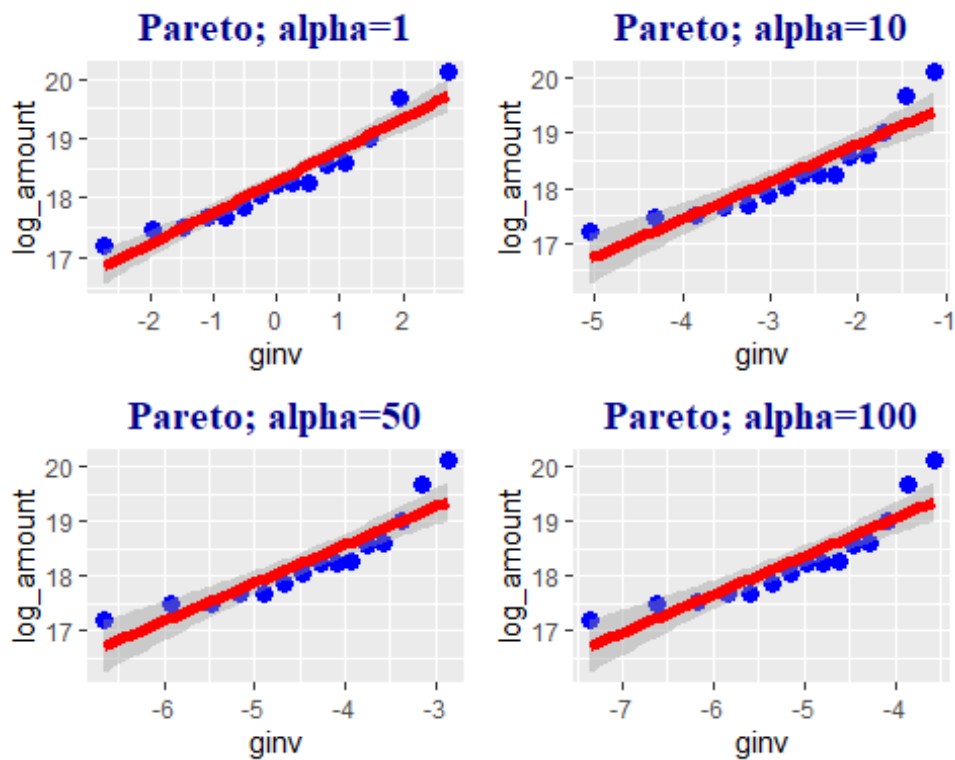
➔ 2)에서 추정한 결과와 마찬가지로 Pareto분포의  $R^2$ 값이 가장 높았다. Pareto분포를 이용하여 절단된 자료의 5만원 이하 사고 건수  $n_1$ 을 추정하였을 때, 실제 값 1039와 가장 유사하였다. 모수 추정 결과, 기존의 값과 큰 차이를 보이지 않았다.

## <Part 2> 특수건물(아파트) 화재보험료 계산

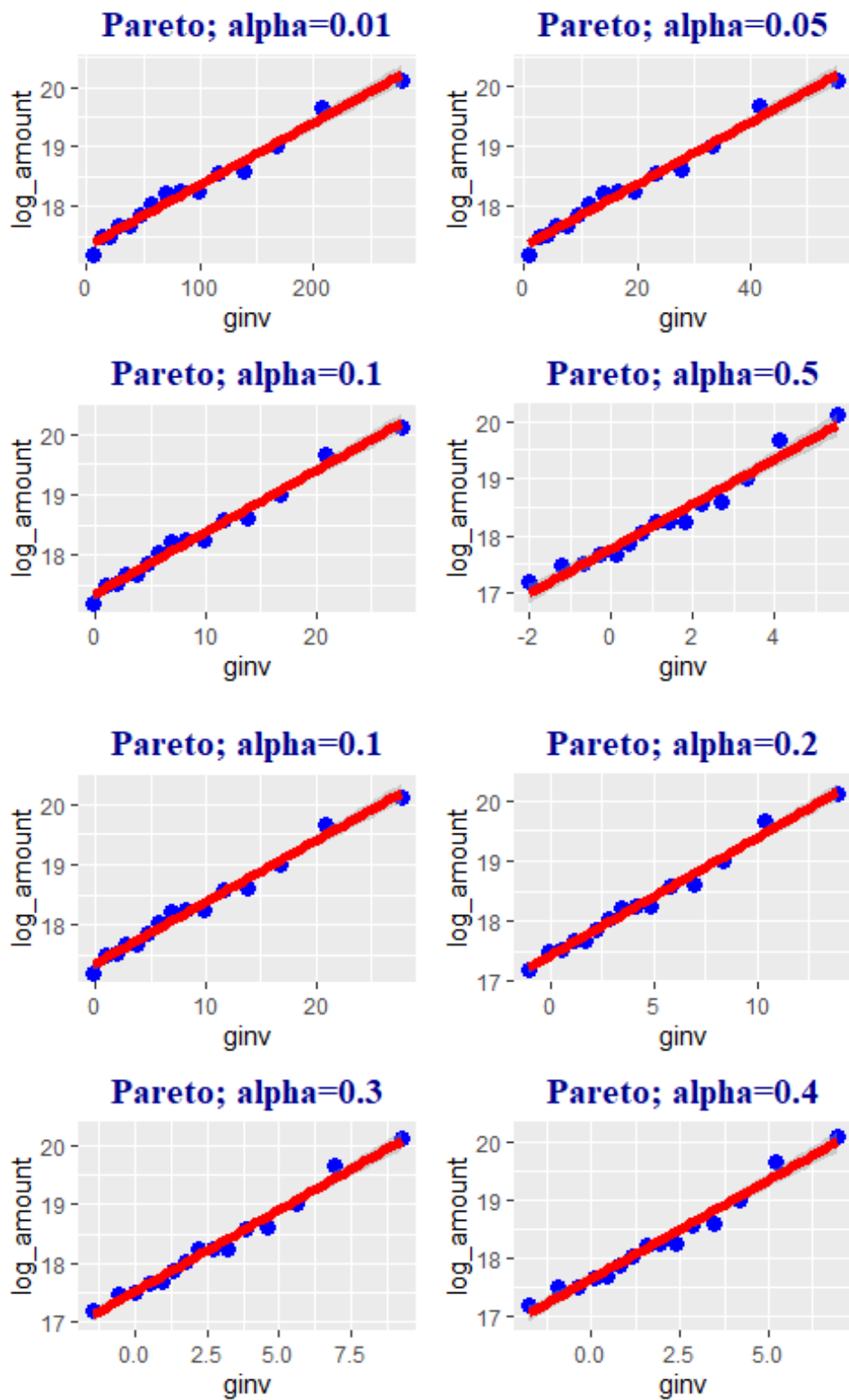
1. 2010년 업종별 피해 건수/금액 규모별 현황자료 중 아파트자료를 이용하여 화재 피해액(X)를 나타내는 위의 5개 후보분포에 대해 Q-Q plot 을 각각 그려보시오.

Pareto 분포의 경우 alpha값을 바꿔가며 시행하였다.

# alpha = 1, 10, 50, 100

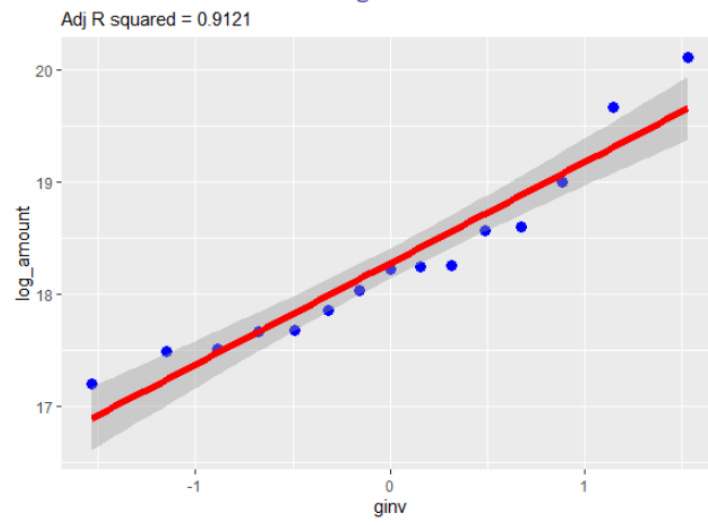


Alpha가 1일 때 제일 직선에 가까웠으므로 alpha가 1~4일 때의 qq plot을 그려보았다.

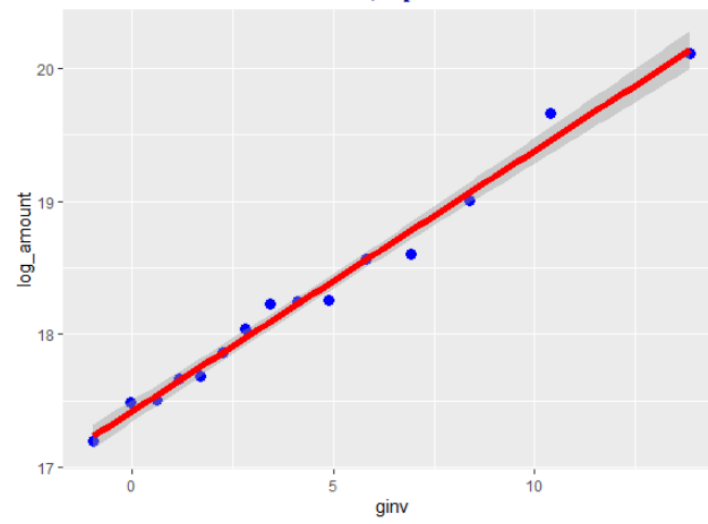


Alpha=0.2일 때 가장 Q-Q Plot이 직선에 가까우므로,  $\alpha=0.2$ 인 pareto 모델과 다른 모델들을 비교해보도록 한다.

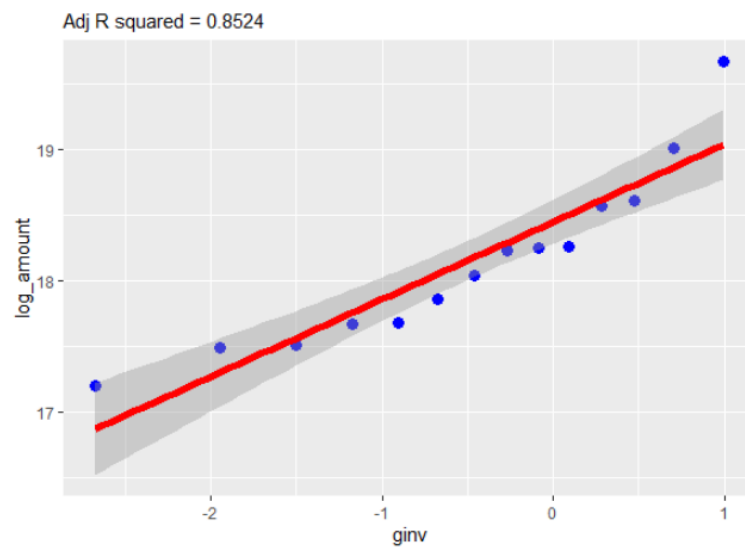
### Log Normal



### Pareto; $\alpha=0.2$

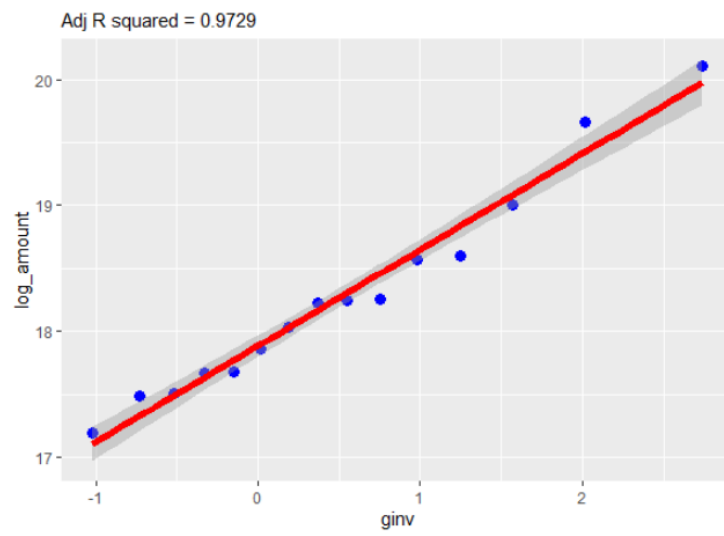


### Weibull

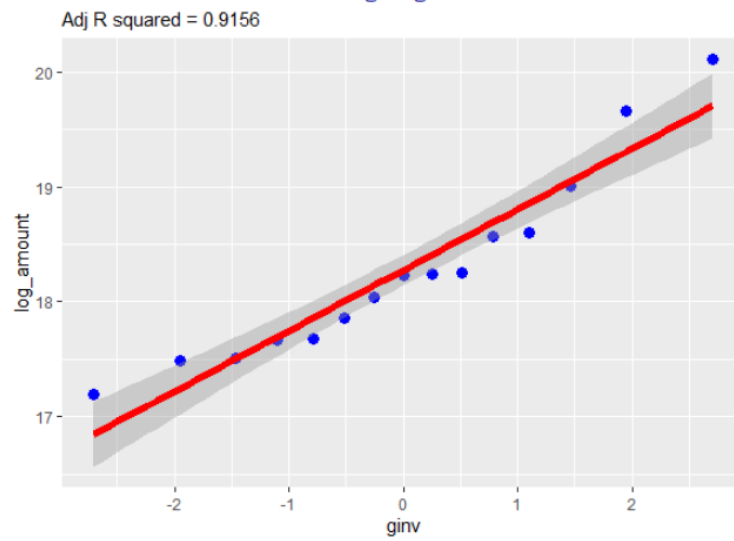




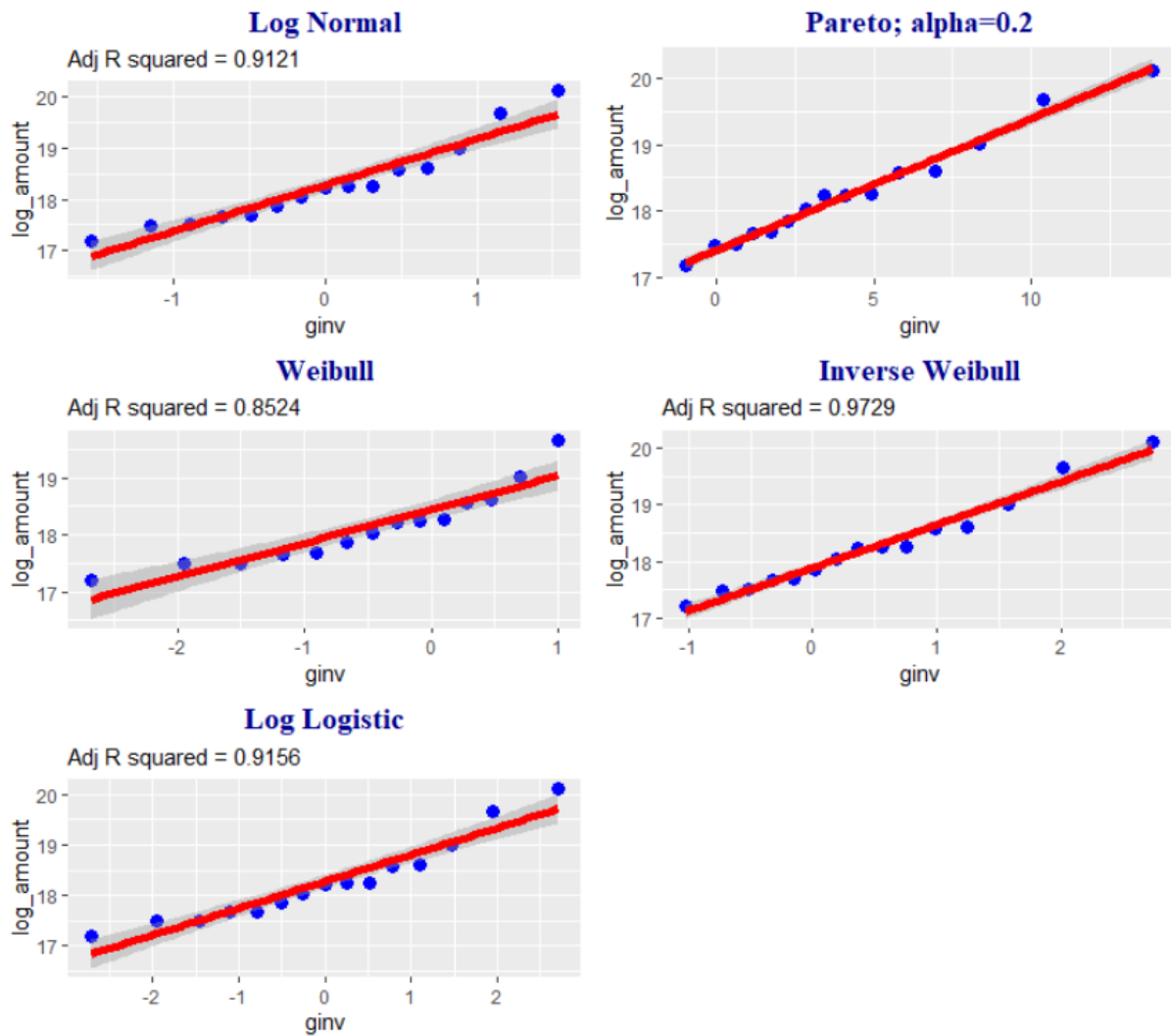
### Inverse Weibull



### Log Logistic



2. 위 Q-Q plot 을 근거로 피해금액(X)의 가장 적절한 확률분포 2개를 찾아보시오.



➔ 위에 5가지 모형의 qqplot을 비교해본 결과 직선에 가장 가까우면서 adjusted R squared가 제일 높은 Inverse Weibull 분포와 그 다음으로 adjusted R squared값이 높은 Pareto모델과 Log Logistic 모델을 최적의 분포로 선택한다.

3. 위에서 선택된 2가지 최적모형을 이용하여 자기부담금 (Deductible)을 A=0, 100, 500 만원, 보상한도 (Limit)를 B= 1000만, 2000만, 5000만, 1억, 2억 으로 각각 조정하였을 때 적정 보험료 (premium)를 각각 계산하고 차이를 서로 비교해 보시오.

공제금액(A)	실제보상한도(B)	보험료(P1) Inverse Weibull	보험료(P2) Log Logistic	보험료(P3) Pareto
0	1000만원	115만원	115만 3천원	115 만 3 천원
	2000만원	225만 7천원	230만 5천원	230 만 5 천원
	5000만원	484만 1천원	557만 3천원	557 만 4 천원
	1억원	704만 3천원	910만 9천원	909 만 8 천원
	2억원	843만 4천원	1110만 6천원	1109 만 4 천원
100만원	1000만원	114만 9천원	115만 3천원	115 만 3 천원
	2000만원	224만 7천원	230만 5천원	230 만 5 천원
	5000만원	479만 2천원	555만 4천원	555 만 6 천원
	1억원	695만 5천원	904만원	902 만 7 천원
	2억원	832만 6천원	1099만 7천원	1098 만 5 천원
500만원	1000만원	113만 7천원	115만 3천원	115 만 3 천원
	2000만원	219만 7천원	230만 2천원	230 만 2 천원
	5000만원	458만 4천원	547만 2천원	547 만 4 천원
	1억원	659만 7천원	875만 2천원	873 만 7 천원
	2억원	788만 9천원	1055만 8천원	1054 만 8 천원

➔ 표를 통해 공제금액이 클수록, 같은 실제보상한도일 때 보험료 P1, P2, P3가 작아지는 것을 알 수 있다. 또한 실제 보상한도가 커질수록 보험료 P1과 P2, P3와의 차이가 커지는데 이는 추정된 모수값에 차이가 있기 때문이다. Inverse Weibull 분포가 Log Logistic 분포에 비해 더 직선에 가까운 최적의 모델이므로 Inverse Weibull 분포 관점에서 분석해보면, Log Logistic 분포로 추정한 P2와 Pareto 분포로 추정한 P3는 P1에 비해 과대추정된 경향이 있다고 할 수 있다.

4. 자기부담금이 없고( $A=0$ ) 보상한도가 없는( $B=\infty$ ) 보험의 적정보험료를 아래 2가지 방법으로 계산하시오.

**방법 a)** 보험료 (premium)의 95% 예측상한 (upper prediction limit)

2010 계약건수 = 5351

→  **$P_a = 142$ 만 1천원**

**방법 b)** 2010년도 경험자료를 이용한 단순 추정 보험료;

보험료 = (2010년총손해액) / (2010년 총대상건수)

= 15476000000 / 5777

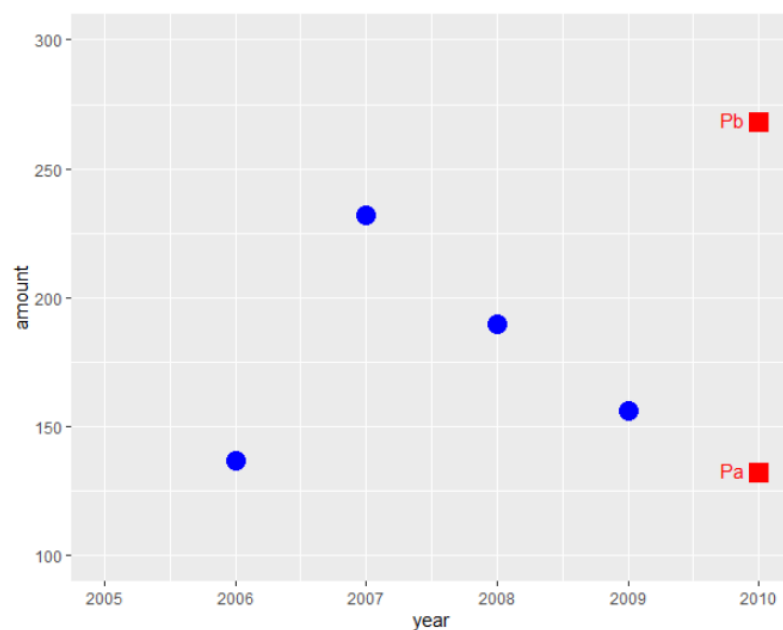
= 2678899

→  **$P_b = 267$ 만 9천원**

5.

가) 위에서 a), b) 두 가지 방법으로 각각 구한 보험료(premium)  $P_a$ ,  $P_b$  를 구하고 이를 <표2-10> (최근 5년간 (2005-2009) 특수건물 주택물건 (16층이상아파트 및 주상복합건물) 화재보험 손해자료)의 평균보험료와 상호 비교 검토해보시오.

	2005	2006	2007	2008	2009	2010 $P_a$	2010 $P_b$
보험료	89만	137만	232만	190만	156만	142만	268만



➔ 평균 보험료는 2006년에 증가했다가 2009년까지 점점 감소하는 추세를 보였다. 최근 5년간의 추세를 감안하면 Pb보단, Pa가 더 적합한 추정치임을 알 수 있다.

나) 보험료(premium) Pa, Pb 를 최근 6년간 (2005-2010) 아파트의 화재보험자료 (2005-2010 연도별 특수건물의 화재현황 자료 이용)에 각각 적용하였을 때 연도별 손해율 ((총손해액 / 총경과보험료)\*100)을 각각 구하시오.

$$\text{총보험료} = (\text{Pa or Pb}) * \text{대상건수}$$

$$\text{손해율} = \text{손해액} / \text{총보험료}$$

Pa = 1.42, Pb = 2.68 (단위 : 건, 원, %)						
	2005	2006	2007	2008	2009	2010
대상건수	12010	11512	6597	3307	4451	5777
화재건수	1617	2427	1893	579	782	617
손해액	89억	112억9천만	105억6백만	42억	82억	18억8천만
총보험료 Pa	170억5천만	163억5천만	93억7천만	47억	63억2천만	82억
총보험료 Pb	321억9천만	308억5천만	176억8천만	88억9천만	119억3천만	154억8천만
손해율 Pa	52.18	69.07	112.15	89.69	130	22.95
손해율 Pb	27.6	36.6	59.4	47.4	68.9	12.2
실제손해율	83.3	71.5	68.5	67.0	118.4	206.9

다) 위에서 구한 손해율을 2010년 특수건물의 화재현황의 <표2-10> (최근 5년간 특수건물 주택물건(16층이상아파트 및 주상복합건물) 화재보험 손해자료)의 실제 손해율과 상호 비교하고 이들 방법의 차이 및 장단점을 검토하시오.

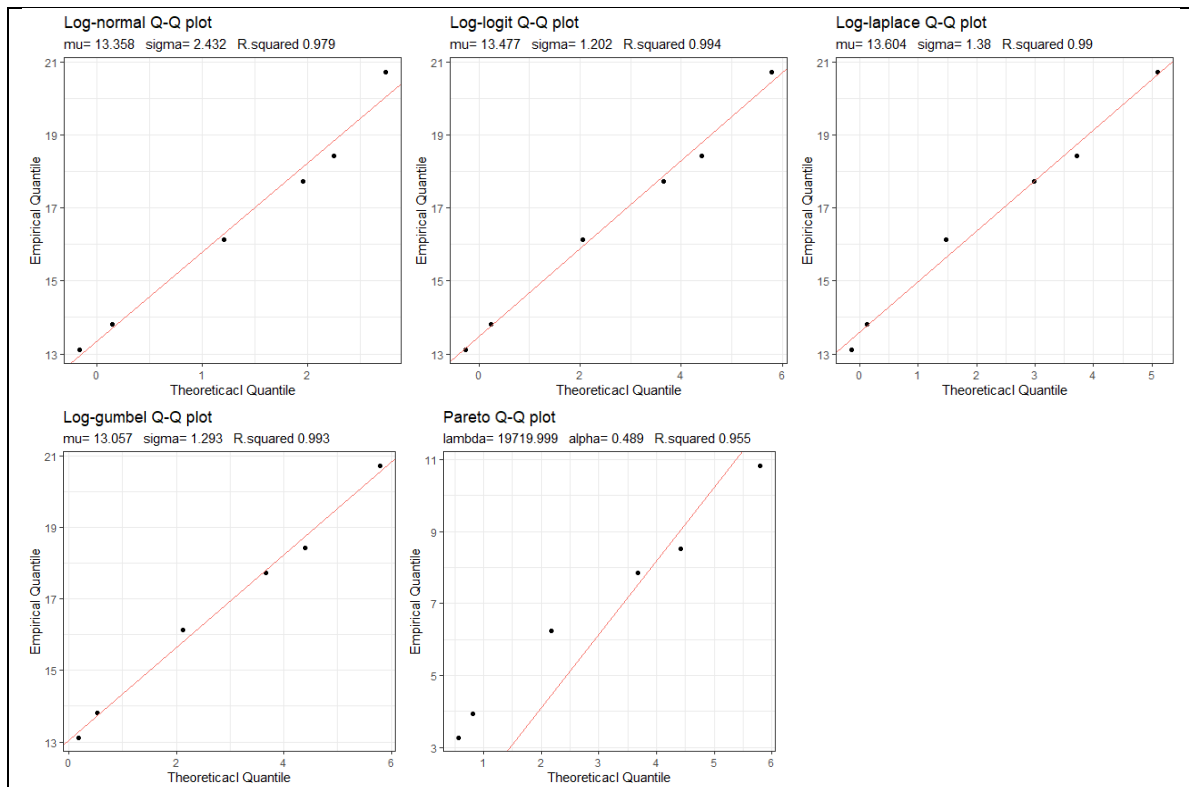
➔ 실제손해율과 추정한 손해율을 비교해봤을 때 Pa로 구한 손해율이 더 실제 손해율과 비슷하다는 것을 볼 수 있다. 그러므로 위 과제에서는 방법a가 더 적합하다.

➔ 방법 a의 경우 보험과 관련된 이론과 원리를 잘 알아야 하는 어려움이 있지만, 덕분에 더 정확한 추정치를 얻을 수 있다는 장점이 있다.

➔ 방법 b의 경우 매우 간단하여 누구나 계산하고 이해할 수 있다는 장점이 있지만, 너무 general 한 공식이므로 추정치의 정확도가 떨어진다는 단점이 있다.

### <Part 3> 학교화재보험료

- a) 피해액별 학교 화재현황통계 자료에서 학교 화재발생 피해액의 Q-Q plot을 그려보고 적합한 분포를 찾으시오.



	모수	$R^2$
Log-normal	(mu, sigma) = (13.358, 2.432)	0.979
Log-logistic	(mu, sigma) = (13.477, 1.202)	0.994
Log-laplace	(mu, sigma) = (13.604, 1.38)	0.99
Log-Gumbel	(mu, sigma) = (13.057, 1.293)	0.993
Pareto	(lambda, alpha) = (19719.999, 0.489)	0.955

- ➔ 2012-2017년에 발생한 학교 화재 피해액의 분포를 구하기 위해, 피해액의 구간을 나누어 QQ plot을 그려보았다.  $R^2$  기준에 의해 Log-logistic 분포와 Log-gumbel 분포가 화재 피해액을 잘 설명하는 분포라 할 수 있다.

- b) 피해액에 대해 log-normal 모형을 적용하여 화재 한 건 당 평균피해액과 적정 학교 화재보험료를 산정하시오.

평균피해액 (평균 사고 심도)	$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ $(\hat{\mu}, \hat{\sigma}) = (13.358, 2.432)$ 평균피해액 = 12164118원
평균화재발생빈도	2013-2017 사고 발생 빈도 = 986 건 2013-2017 유초중등 + 대학 = 105261 개 평균화재발생빈도 = 총 사고 발생 빈도 ÷ 5년 간의 전체 학교 수 = 0.009367192
적정 학교 화재보험료	방법 1) 보상 한도가 없는 경우 $\bar{w}_\alpha$ 의 95% 예측상한의 값을 적정 보험료로 정함 $\bar{w}_\alpha = \hat{\mu}_W + z_\alpha \times \hat{\sigma}_W / \sqrt{n} = 183503$ 원 $\hat{\mu}_W = \text{평균피해액} \times \text{평균화재발생빈도} = 113944$ 원 $\hat{\sigma}_W^2 = 1.898398e + 14$ 방법 2) 경험자료를 이용한 단순 추정 보험료 총손해액/총대상건수 = 105435 원

- c) 2013-2017년 전체 학교수를 근거로 평균 화재발생빈도를 추정하고 a)에서 선택된 최적 모형 2개를 이용해 보상한도가 1억/5억/10억인 경우 적정 학교 화재보험료를 산정하시오.

보상한도	Log-logistic	Log-gumbel
1억	49,051원	46,220원
5억	77,142원	74,832원
10억	91,867원	90,799원

➔ 보상한도가 높을수록 보험료가 높고, Log-logistic 분포로 추정했을 때 보험료가 더 높다.

- d) 위에서 구한 보험료를 2008-2012년도에 적용했을 때 손해율을 구하고 결과의 의미 및 문제점을 검토하고 개선방안을 제시하시오.

분포		Log-logistic			Log-gumbel		
연도	손해액	1억	5억	10억	1억	5억	10억
2008	1,761,628,000	179.848	114.359	96.029	190.864	117.888	97.157
2009	897,342,000	91.155	57.962	48.672	96.738	59.751	49.244
2010	1,397,075,000	141.279	89.834	75.435	149.932	92.606	76.322
2011	818,431,000	81.927	52.094	43.744	86.944	53.702	44.258
2012	1,153,350,000	114.547	72.836	61.162	121.563	75.084	61.881

- ➔ 2013년-2017년 자료를 통해 산정한 보험료를 이용해 과거 2008년-2012년 학교 화재 피해액 자료에 대한 손해율을 계산한 결과는 다음과 같다.
- ➔ 한도가 1억일 경우에는 2009년과 2011년을 제외하고 보험료를 통해 피해액을 감당할 수 없을 만큼 많은 손실이 생겼다.
- ➔ 2013-2017의 화재 사고 중 986건의 화재 중 962건의 화재(97.6%)가 1억 미만의 화재인 것을 함께 고려한다면, 한도를 1억으로 한다면 보험회사의 손실이 클 것으로 예상된다.
- ➔ 한도가 5억원인 경우 2008년을 제외하고는 손실율이 100% 미만이고, 한도가 10억원인 경우에는 과거 5년 간의 손해율이 모두 100% 미만인 것을 확인할 수 있다.
- ➔ 2008년도와 같이 큰 규모의 사고가 발생했을 때를 대비해 보상 한도를 5억 이상으로 잡는 것이 합리적이다. 전반적으로 Log-logistic 분포를 가정할 때, 더 높은 값의 보험료가 산정된다.