

이론통계학2

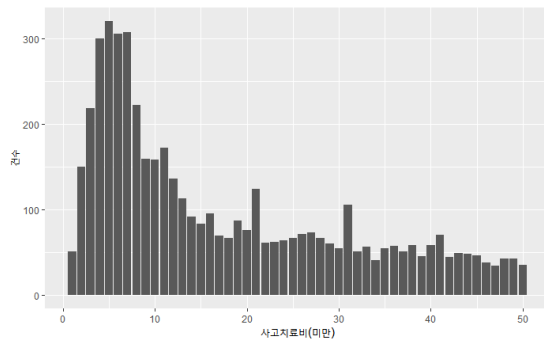
Project #2.

1조

202STG26	박지윤
202STG27	이수현
212STG04	김이현
212STG12	박윤정

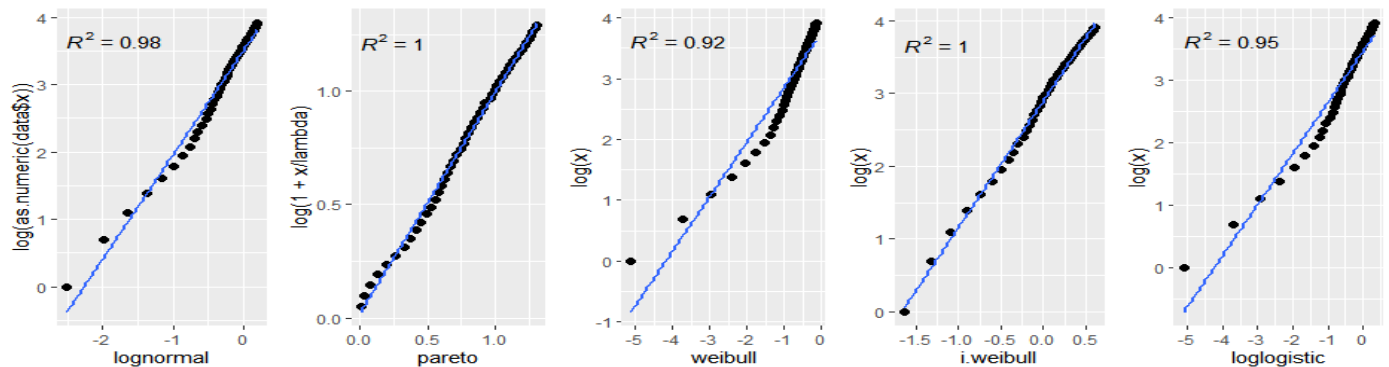
Part 1 : 상해보험

1. 계급화 된 자동차 사고 치료비 (X) 의 histogram을 그려보시오.



자동차 사고 치료비가 4만원 이상, 5만원 미만의 금액의 경우 320건으로 가장 높은 값을 가졌고, 치료비가 커질수록 발생 건수가 줄어드는 오른쪽 꼬리가 긴 모양을 띄는 것을 확인할 수 있다.

2. 위 histogram 및 Q-Q plot 을 이용하여 사고 손실액(X)에 가장 적합한 확률분포 2개를 찾아보시오.



	Log-normal	weibull	frechet	Log-logistic	pareto
R^2	0.981	0.92	0.996	0.95	0.999

Q-Q plot을 확인해봤을 때, pareto 분포와 frechet 분포가 가장 직선에 가까운 형태를 보인다. R^2 값 또한 각각 0.999, 0.996로 가장 높은 값을 가졌다. 따라서 자동차 사고 치료비 분포는 frechet와 pareto 분포를 따른다고 판단했다. 추정된 모수는 frechet에서 $\tau=0.573$, $c=5.281$, pareto에서 $\lambda=19$, $\alpha=0.665$ 이다.

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

자기부담금	보상한도	P1 (frechet)	P2 (pareto)
0	50만원	9,570원	9,540원
	100만원	15,230원	14,980원
	200만원	23,390원	22,380원
	500만원	39,510원	35,800원
	1000만원	57,260원	49,350원
10만원	50만원	8,110원	8,110원
	100만원	13,440원	13,190원
	200만원	21,340원	20,300원
	500만원	37,190원	33,460원
	1000만원	54,810원	46,890원

20만원	50만원	7,200원	7,150원
	100만원	12,250원	11,920원
	200만원	19,900원	18,760원
	500만원	35,510원	31,670원
	1000만원	53,000원	44,980원

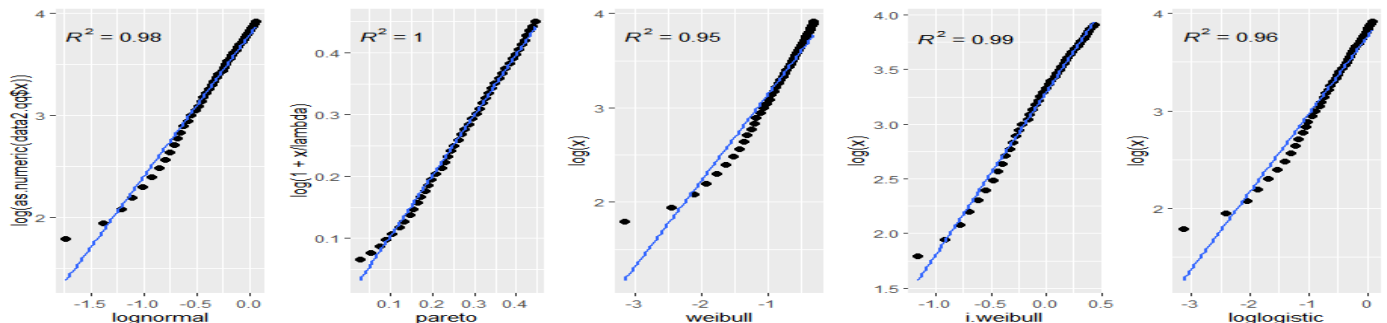
Frechet과 Pareto 분포를 이용해 구한 적정보험료이다. 모든 경우에서 Frechet 분포를 이용해 계산한 보험료가 약간 더 높으나 비슷한 값으로 측정되었다. 또한 공제가 없고 보상한도가 50만원인 경우 기존의 보험료인 3130원보다 높은 값이 적정보험료로 계산되었다.

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

	P1 (frechet)	P2 (pareto)
손해율(%)	100.023	100.337

공제가 없고 보상한도가 50만원이었을 때의 손해율이다. 보험료가 높아진 만큼 두 경우 모두 기존의 손해율인 305.8%보다 손해율이 낮아졌다. 그리고 두 경우 모두 손해율이 100%에 가까움으로 적절히 책정된 보험료라고 할 수 있다.

5. 만약 자료에서 5만원의 자기 부담금이 있어 5만원 이하의 사고자료가 없는 경우 아래의 절단된 (truncated)자료 (아래 표 3)를 가지고 손해액()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 모수 추정값과 서로 비교해보시오.



	Log-normal	weibull	frechet	Log-logistic	pareto
R^2	0.983	0.948	0.993	0.968	0.999

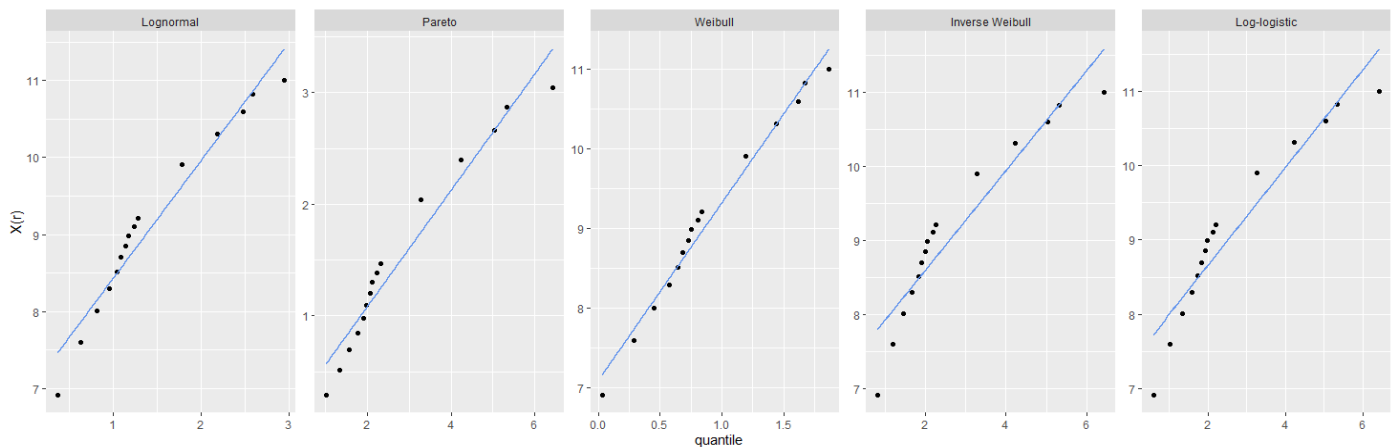
절단된 자료에 대해 Q-Q plot을 그렸을 때 pareto 분포가 가장 직선 형태에 가까움으로 이를 손해액의 분포로 결정한다. 이때의 모수는 $\lambda = 93.002$, $\alpha = 1.414$ 로 두 모수 모두 자기부담금이 없는 경우에서 추정한 모수 값보다 크다.

	m	n
추정값	8498	1584
상대오차(%)	0.651	52.454

이후 missing 자료가 있는 다항분포를 사용하여 MLE 방법으로 총사고 건수 m을 추정했을 때 m=8498, 5만원 이하의 사고건수 n1=1584로 추정되었다. 실제 m=8443, n1=1039와의 상대오차는 -0.651%, 52.454%이다.

Part 2 : 특수건물(아파트) 화재보험료 계산

- 2010년 특수건물 화재현황에서 2010년 업종별 피해 건수/금액 규모별 현황자료 중 아파트 자료를 이용하여 화재 피해액을 나타내는 Q-Q plot 그려보시오.



	Log-normal	weibull	frechet	Log-logistic	pareto
R^2	0.950	0.982	0.892	0.904	0.989

- Q-Q plot을 근거로 피해액의 가장 적절한 확률분포 2개를 찾아보시오.

위의 Q-Q plot에서 5가지 모형 모두 직선에 가까운 형태를 보인다. 그 중 Weibull모형이 가장 직선에 가깝게 관측치들이 존재한다. 이를 제외하고는 모두 비슷한 수준을 보이기에 5가지 모형에 대해 R^2 를 계산해보았다. 그 결과 Pareto분포가 0.984로 가장 높았고 그 다음이 0.956으로 Weibull 분포이다. 따라서 아파트의 화재 피해액을 가장 잘 나타내는 모형은 Weibull과 Pareto분포이다. Weibull모형을 이용하여 모수를 추정한 결과 $\mu = 7.098$, $\sigma = 2.233$ 이다. Pareto모형을 이용하여 추정한 결과 $\lambda = 3000$, $\alpha = 1.882$ 이다.

- 2가지 최적모형을 이용하여 자기부담금과 보상한도가 아래와 같을 때 적정 보험료를 계산하시오.

자기부담금	보상한도	P1 (Weibull)	P2 (Pareto)
0원	1000만원	217,000원	263,600원
	2000만원	268,200원	303,100원
	5000만원	309,600원	334,500원
	1억원	320,800원	347,300원
	2억원	323,400원	354,500원
100만원	1000만원	166,700원	188,500원
	2000만원	213,300원	223,900원
	5000만원	252,300원	253,500원
	1억원	263,000원	266,000원
	2억원	265,500원	273,200원
500만원	1000만원	88,700원	78,200원
	2000만원	121,700원	102,300원
	5000만원	152,300원	126,300원
	1억원	161,400원	137,600원
	2억원	163,600원	144,400원

자기부담금은 커질수록, 보상한도는 작아질수록 보험료는 감소한다. 자기부담금=0, 보상한도=1000만원, 2000만원일 때 Weibull모형을 이용한 경우에 비해 Pareto 모형을 이용한 경우 보험료가 더 비쌌으나 자기부담금과 보상한도가 커질수록 Pareto 모형을 이용한 경우가 보험료가 더 저렴했다.

4. 자기부담금은 없고 보상한도가 없는 보험의 적정보험료를 2가지 방법으로 계산하시오.

아래 두가지 방법으로 자기부담금은 없고 보상한도가 없는 보험의 적정보험료를 구하였다. Pa를 Weibull, Pareto분포를 이용하여 구한 결과 각각 323,700원, 363,400원이다. 두번째 방법을 이용한 Pb는 326,000원이다.

$$E(W) = E(N) \cdot E(Y) \quad \dots\dots Pa$$

$$W = (\text{해당 연도의 총손해액}) / (\text{해당 연도의 총대상건수}) \quad \dots\dots Pb$$

5. 2010년 특수건물 화재현황에서 <표 2-10> (2005-2009년 특수건물 중 주택건물 화재보험 손해자료)를 이용하여 계산하시오.

1) 5년간의 보험료를 구하고 실제 평균보험료와 비교하시오.

연도	대상건수	화재건수	손해액 (백만원)	총보험료 (천원)			
				(a)-weibull	(a)-pareto	(b)	실제
2005	12010	1617	88,990	408	458	326	8,900
2006	11512	2427	112,910	639	717	326	13,700
2007	6597	1893	105,060	870	976	326	23,200
2008	3307	579	42,120	531	596	326	19,000
2009	4451	782	82,160	532	598	326	15,600
2010	5777	617	1883.5	324	363	326	•

위에서

추정한 Weibull과 pareto분포의 모수를 이용하여 2005년~2009년의 보험료를 구하였다. 두 모형에 이용한 Pa는 실제 평균보험료와 큰 차이가 크며 보험료가 실제 값보다 작게 추정되었다. 이에 반면에 Pb의 경우 실제 총손해액과 총대상건수를 이용하여 구하였기에 실제 평균보험료와 근접한 값을 보인다. 모수를 추정할 때 사용한 데이터는 2010년 아파트 화재 자료인 반면 2005-2009년 자료는 특수건물 중 주택건물(16층이상 아파트 및 주상복합건물) 화재 자료이기에 차이가 큰 것으로 생각된다.

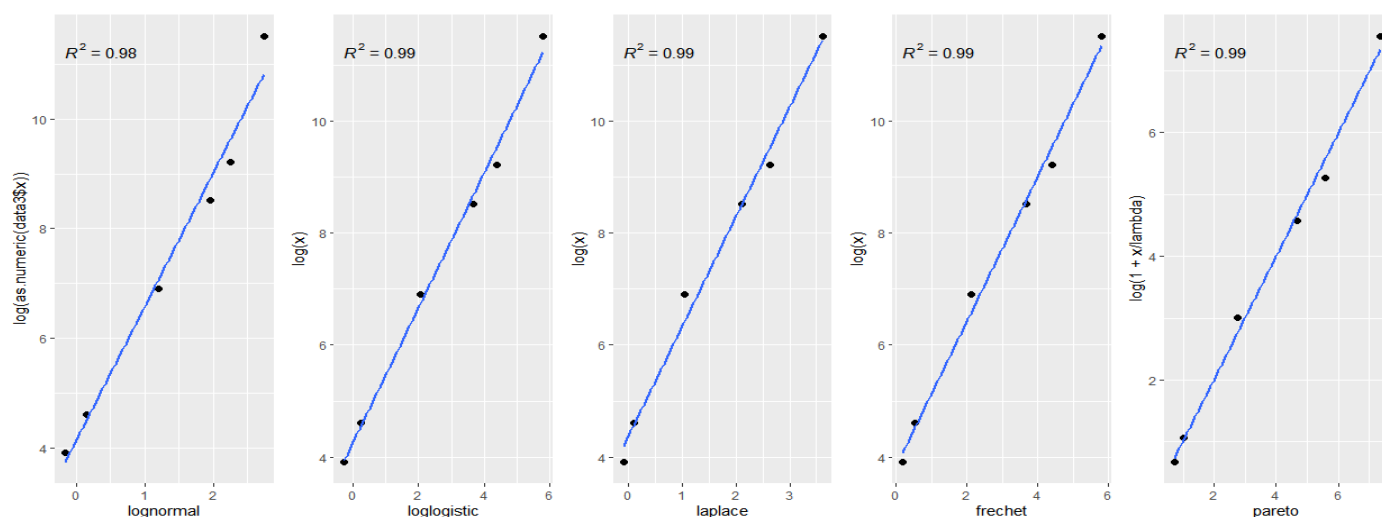
2) 보험료를 6년간(2005-2010) 아파트의 화재보험자료에 적용하였을 때 연도별 손해율을 구하고 실제 손해율과 비교하시오.

연도	대상건수	화재건수	손해액 (백만원)	손해율 (%)			
				(a)-weibull	(a)-pareto	(b)	실제
2005	12010	1617	88990	1816	1617.6	2272.7	83.3
2006	11512	2427	112910	1535.1	1367.4	3008.3	71.5
2007	6597	1893	105060	1831.3	1631.2	4884.6	68.5
2008	3307	579	42120	2400.4	2138.1	3906.5	67
2009	4451	782	82160	3466.8	3088	5661.6	118.4
2010	5777	617	1883.5	100.7	89.7	100	•

위에서 언급한바와 같이 2005~2009년의 Pa를 적용한 경우 오차가 크므로 손해율이 엄청나다. Pb를 이용한 경우 항상 손해율은 100%로 받은 보험료와 보상금액이 동일하다. 하지만 실제 손해율의 경우 2009년을 제외하고는 100%보다 작으므로 보험 회사의 수익으로 이어졌음을 알 수 있다.

Part 3 : 학교 화재 발생 자료

1. 학교 화재발생 피해액의 Group화된 자료의 Q-Q Plot 및 모수 추정



	log-Normal	Log-logistic	Log-laplace	Frenchet	Pareto
R- squared	0.979	0.994	0.990	0.993	0.998

Q-Q plot을 그려봤을 때 Log-logistic과 pareto가 가장 직선에 가까움으로 최적 분포로 선택한다.

Log-logistic distribution		Pareto distribution	
lamda	alpha	lambda	alpha
4.147	2.432	52	0.788

Q-Q Plot으로 모수를 추정한 결과 log-logistic 분포의 모수는 $\lambda = 4.147$, $\alpha = 2.432$, pareto 분포의 모수는 $\lambda = 52$, $\alpha = 0.788$ 로 도출되었다.

2. 학교화재의 피해액에 대하여 log-normal 모형을 적용하여 화재 한건 당 평균피해액 및 적정 학교 화재보험료를 산정하시오.

보험료 = 평균 사고 빈도 * 평균 사고 심도

평균 사고 빈도 = 5년간 총 사고 건수 / (5 * 전체 학교 수)

한도	평균 사고 빈도	평균 사고 심도	보험료
Infinite	0.009	1217.206	115,600원
10억원	0.009	1004.846	95,500원

교육통계서비스에서 제공한 자료에 따르면 현재 우리나라의 총 학교수는 20761개이다. 추정 정확도를 높이기 위해 5년치의 데이터를 사용했을 때 총 사고 986건과 전체 학교 수 103,805 개를 이용하여 구한 평균 사고 빈도 $E(N)$ 은 0.009이다. 데이터의 한도를 확신할 수 없어 한도가 없는 경우, 한도가 10억인 경우에 대해 각각 보험료를 계산했다. R integrate 함수와 log-Normal 분포의 확률밀도함수를 이용하여 구한 평균 사고 심도는 한도가 없을 때 1004.846, 한도가 10억일 때 1217.206이다. 두 결과를 통해 산정한 적정 보험료는 각각 95,500원, 115,600원이다.

3. 2013-2017년 전체 학교수를 근거로 평균 화재발생빈도를 추정하고 a)에서 선택된 최 적 모형 2개를 이용하여 보상 한도가 1억/ 5억/10억 인 경우 적정 학교 화재보험료를 각각 산정 하시오.

위 문제에서 선택된 최적 모형인 log-Normal 모형과 Pareto 모형을 통해 보상 한도가 1억, 5억, 10억인 경우에 대한 적정 화재보험료를 산정한다.

2013-2017년 전체 학교 수를 근거로 구한 평균 화재 발생 빈도 $E(N)$ 은 0.009이다.

각 보상 한도에 따른 평균 사고 심도 및 적정 보험료는 다음과 같다.

보상한도	모형	평균 사고 심도 $E(Y)$	보험료
1억원	log-logistic	523.651	49,740원
	Pareto	503.494	47,820원
5억원	log-logistic	823.531	78,220원
	Pareto	807.047	76,660원
10억원	log-logistic	980.727	93,160원
	Pareto	973.469	92,470원

보상 한도가 높을수록 평균 사고 심도와 보험료가 높아지는 것을 알 수 있다. 또한, 대체적으로 Pareto 모형보다 log-logistic 모형을 이용할 경우 보험료가 높게 측정된다.

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

손해율 = 총 손해액 / (가입 건수 * 보험료)

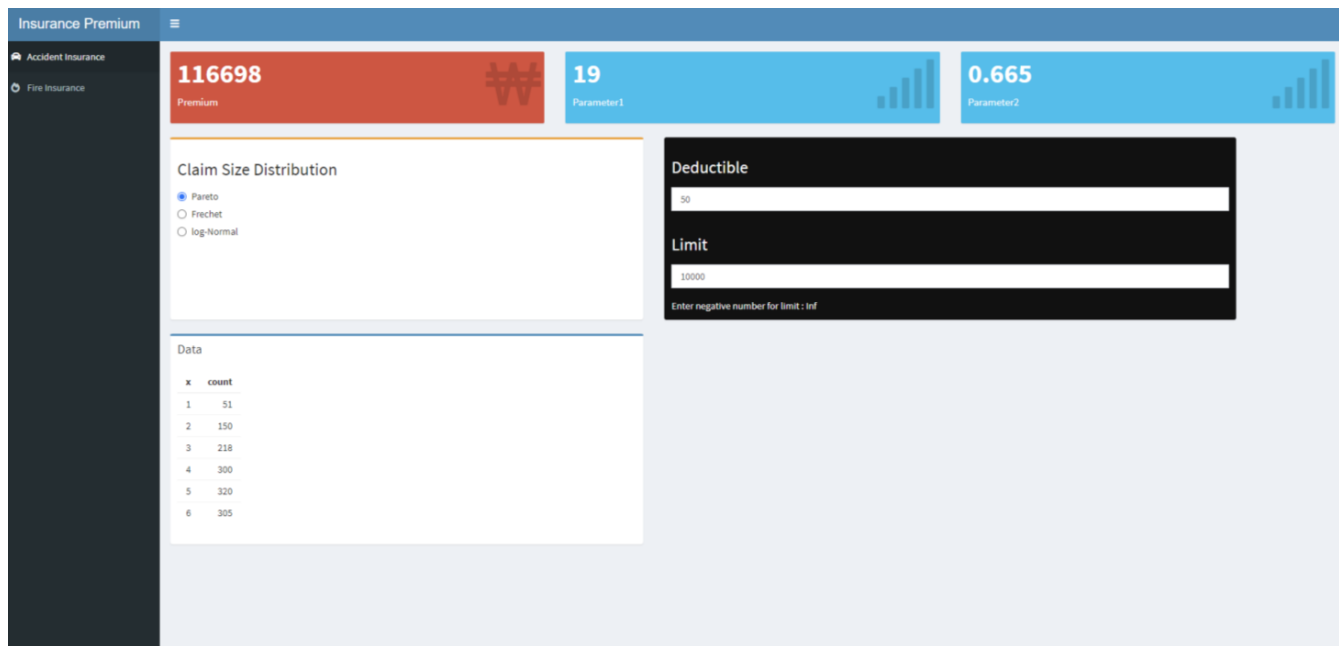
앞서 계산한 각 모형과 보상 한도에 따른 보험료를 2008-2012년도에 적용하여 구한 손해율은 다음과 같다.

연도	보상한도(원)	Log-logistic	Pareto
		손해율(%)	손해율(%)
2008	1억	179.056	186.224
	5억	113.854	116.180
	10억	95.605	96.316
2009	1억	90.689	94.320
	5억	57.666	58.844
	10억	48.423	48.783
2010	1억	140.517	146.143
	5억	89.349	91.174
	10억	75.028	75.586
2011	1억	80.793	84.028
	5억	51.373	52.423
	10억	43.139	43.460
2012	1억	112.963	117.485
	5억	71.829	73.296
	10억	60.315	60.764

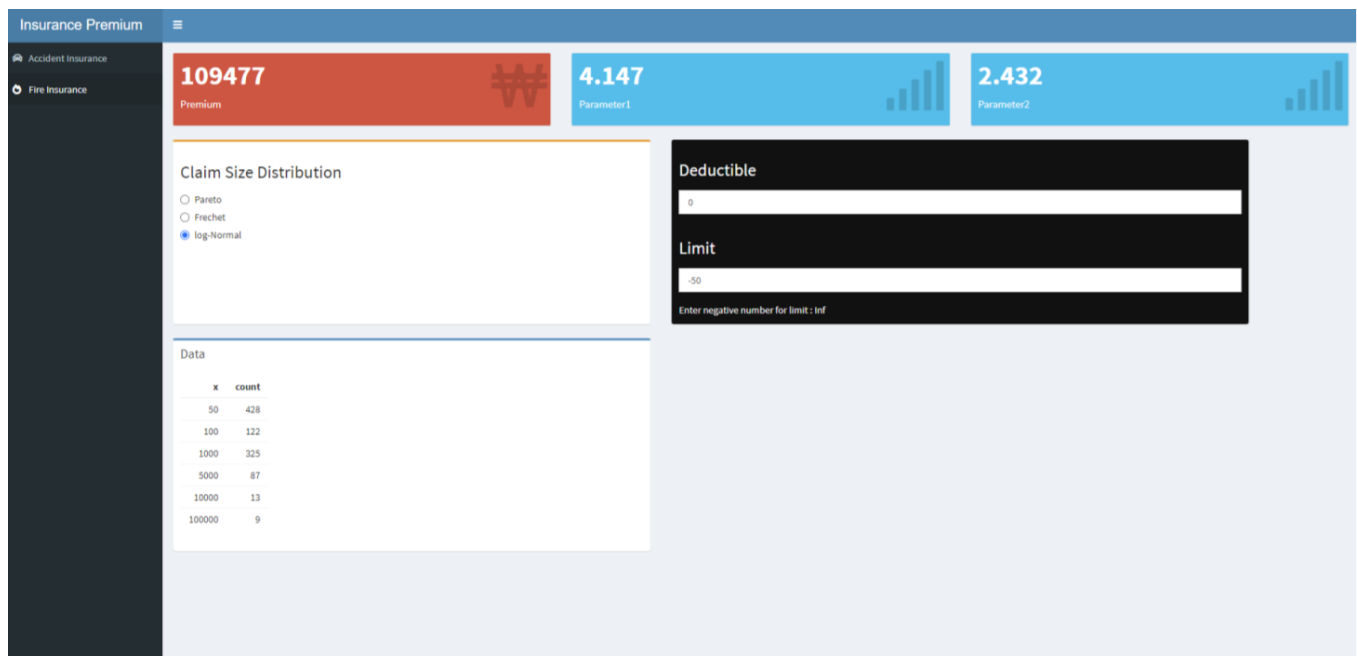
해당 결과는 2013년~2017년 자료에서 추정된 보험료를 2008년~2012년 자료에 적용한 것으로, 화재 발생 횟수 차이로 인해 변동이 큰 것으로 보인다. 모형을 비교했을 때, Pareto 모형이 log-logistic 모형보다 상대적으로 높은 손해율을 보인다. 적절하게 책정된 보험료의 손해율은 100(%)이다. 두 모형 모두 한도가 1억원일 때는 대부분 손해율이 100을 초과한다. 이는 보험회사가 굉장한 손해를 보고 있음을 뜻한다. 반대로 한도가 높아질수록 적정 보험료는 높게 책정되고 그에 따라 보험회사의 손해율은 낮아진다. 특히 한도가 10억에 가까워질수록 손해율이 굉장히 작아지는데 이는 고객이 지불하는 보험료에 비해 적절한 보험 혜택을 받지 못했음을 뜻한다. 이를 통해 보험 회사와 고객이 모두 만족할 수 있는 한도를 책정하는 것이 중요함을 알 수 있다. 또한 2008년에서 한도가 5억원일 때, 2012년에서 한도가 1억원일 때 손해율이 100보다 조금 더 크게 나왔는데 이때 적절한 자기부담금을 책정한다면 손해율을 100에 가깝게 낮출 수 있을 것으로 예상된다.

Part 4 : 손해보험료 계산 Application 개발

<https://jiyoon-ing.shinyapps.io/project2/>



Ex) 자기부담금 50만원, 한도 10억일 때의 파레토 모형 상해 보험료



Ex) 자기부담금 0원, 한도로 음수 값을 넣었을 때의 로그노말 모형 화재 보험료

(한도가 없는 보험료를 계산해 줌)