MT7027: Project 1

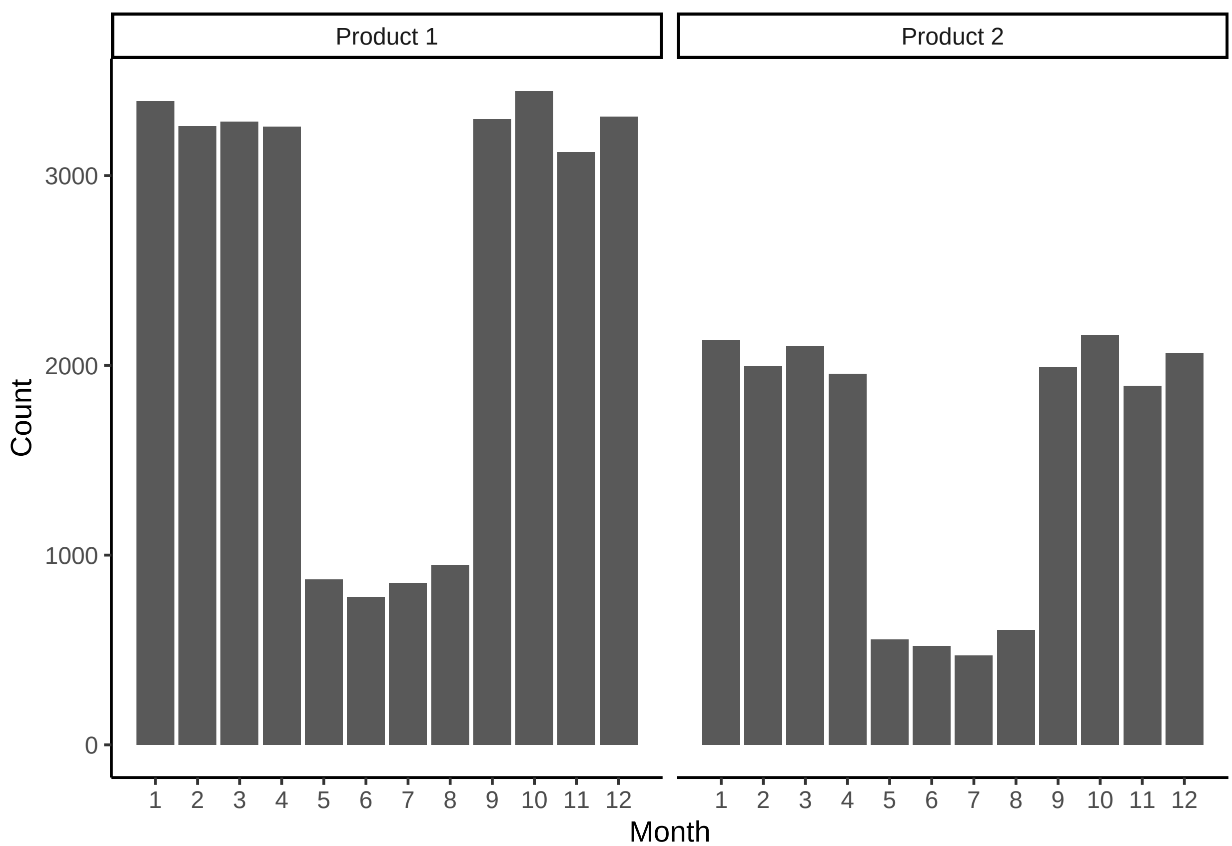
Anton Stråhle, Jan Alexandersson & Max Sjödin

Introduction

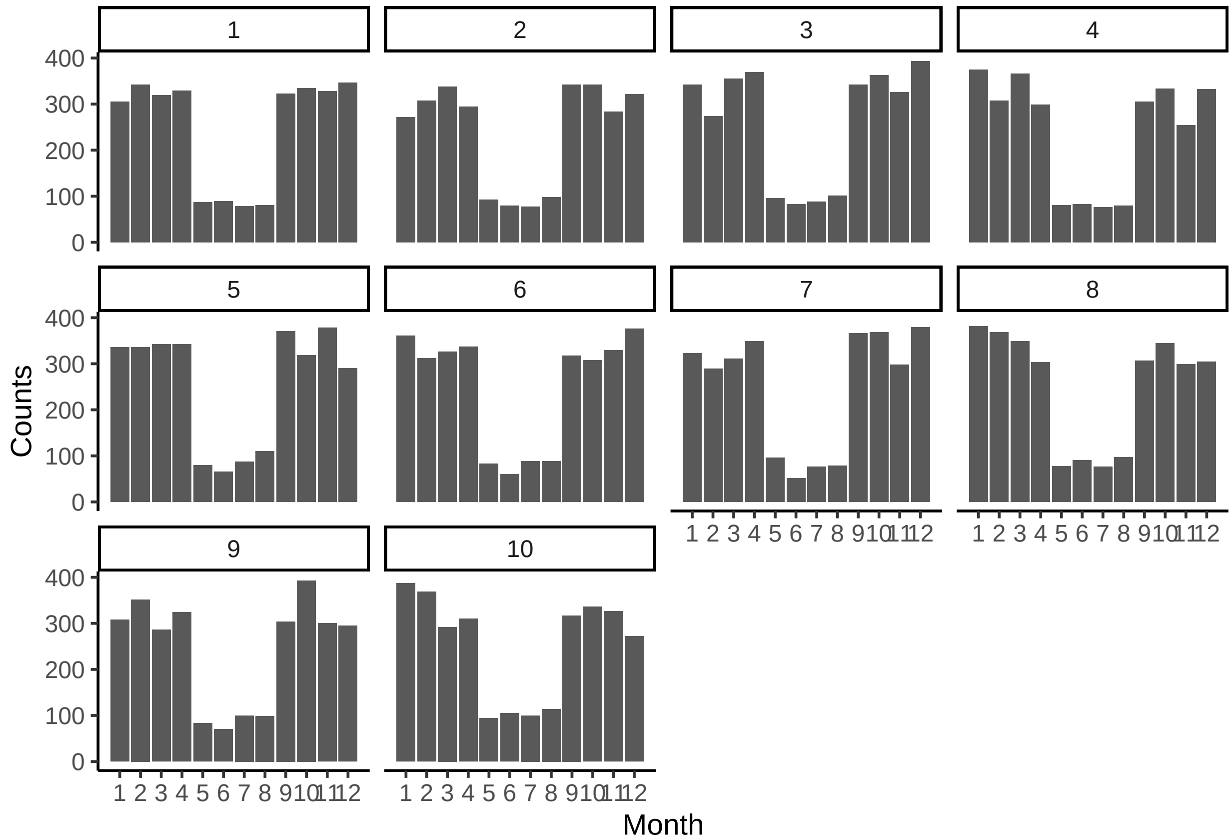
In this project we are dealing with data concerning two different insurance branches collected over 10 years. We do not know anything about the sizes of the two insurance portfolios except the fact that their size has not changed over the decade which the data spans. Furthermore, the insurance products are of the non-life type and are paid out in lump payments.

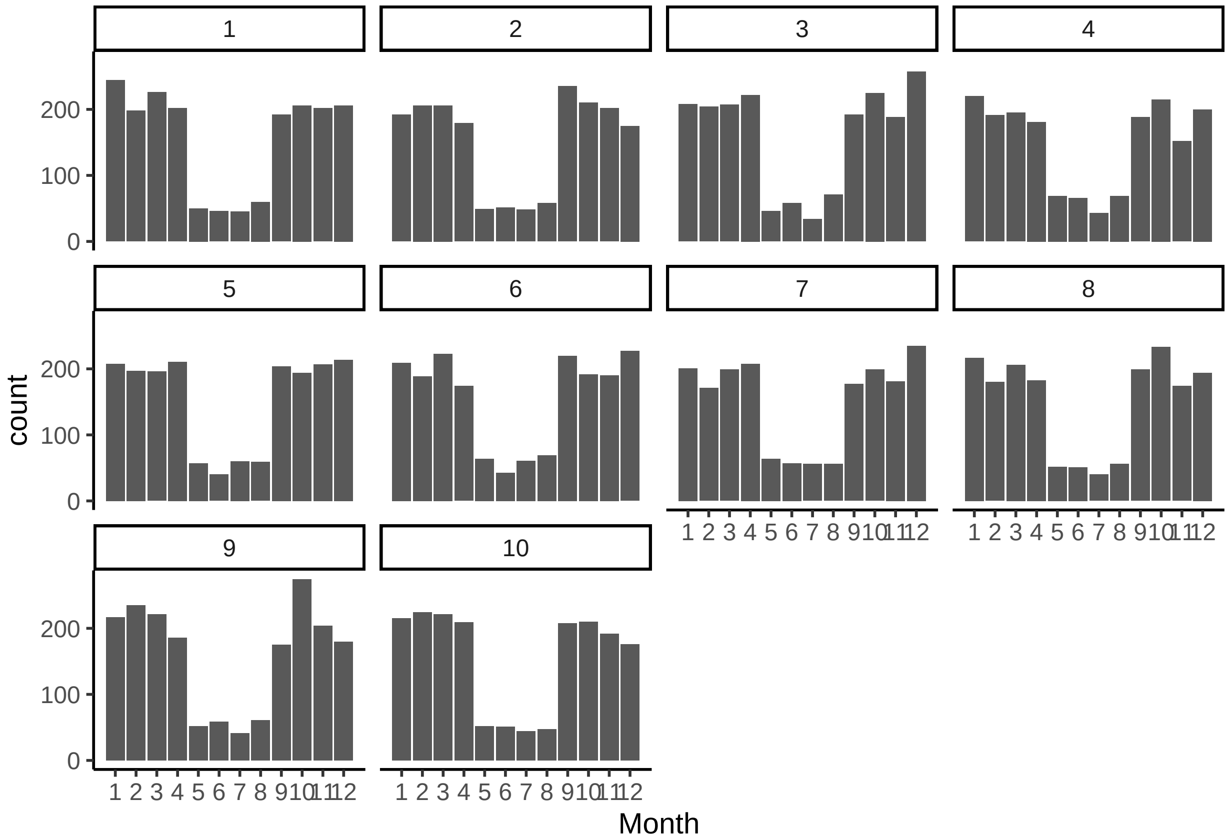
Exercise 1

In this exercsie we want to find trends in the data for the two insurance branches in order to model future claims in a block-wise manner. Since the data is structured in a way such that we only have the claim day (i.e. $1, 2, ..., 3650$) we assume that 365 day/year and that a month is 365/12 days (to get 12 months).



From the previous histograms it seems reasonable to divide the months into two homogenous groups with their own claim processes. One group for months $(1-4, 9-12)$ and one for months $(5-8)$. We also wish to examine if we have homogeneity between the different years during which the claim data was collected.



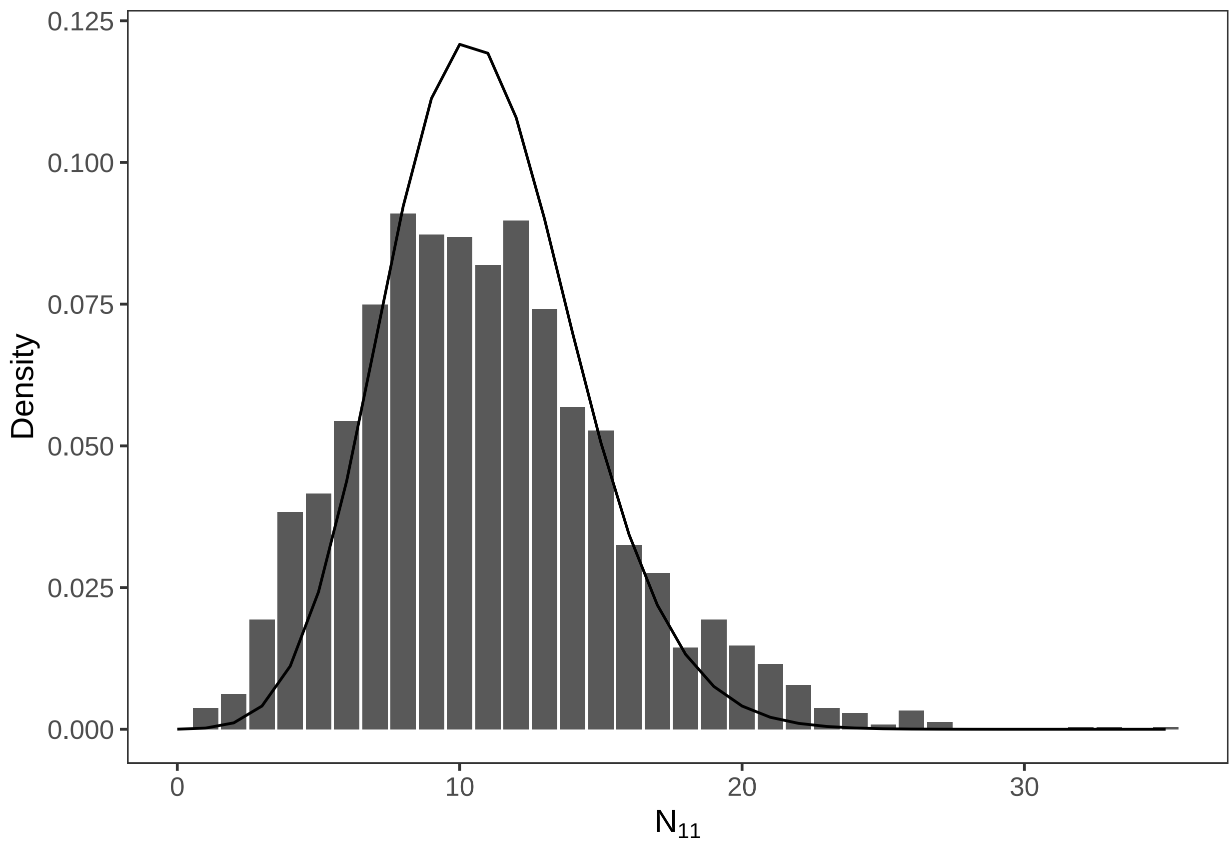


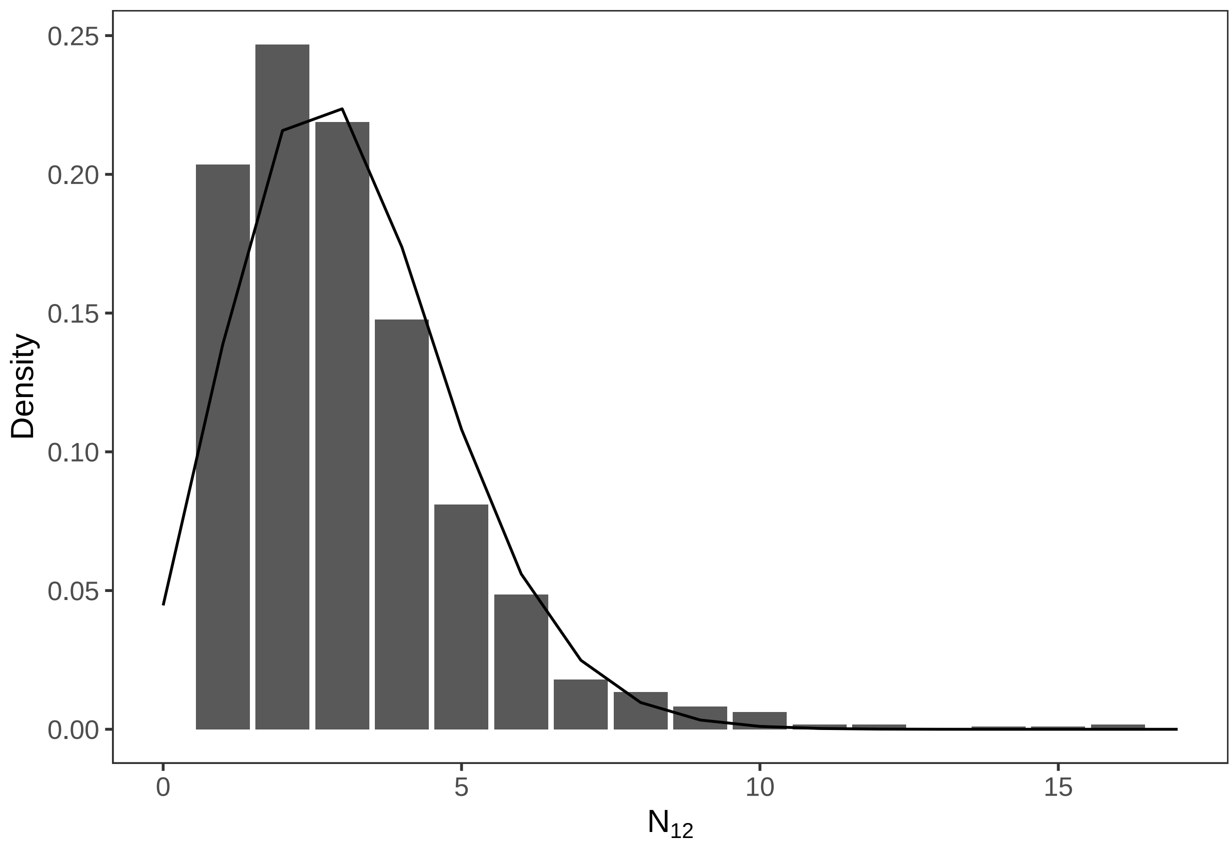

```
## $y
## [1] "Counts"
##
## attr(,"class")
## [1] "labels"
```

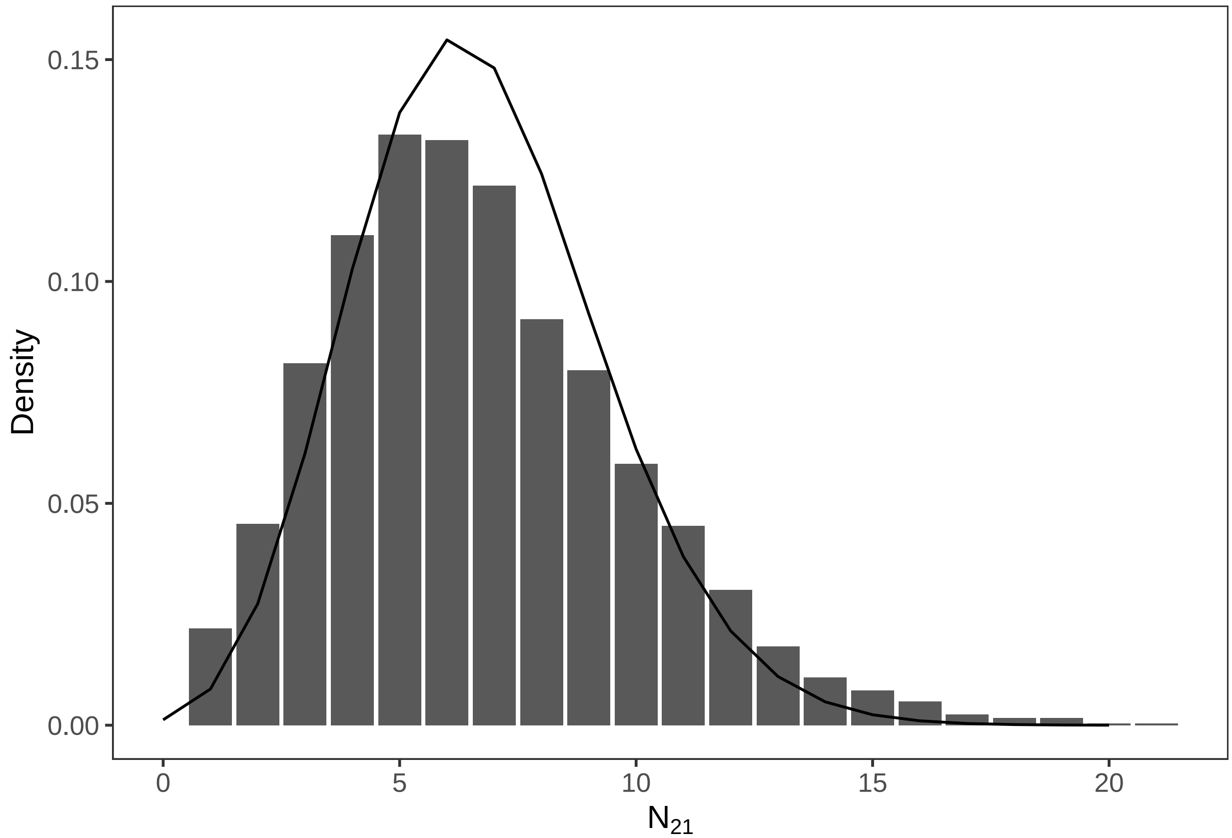
We note from the previous plots that there does not seem to be any major difference in the number of claims between the years.

We now wish to fit homogeneous Poisson distributions Nij (where i represents the insurance type and j represents the season) to the months of each season and each insurance product.

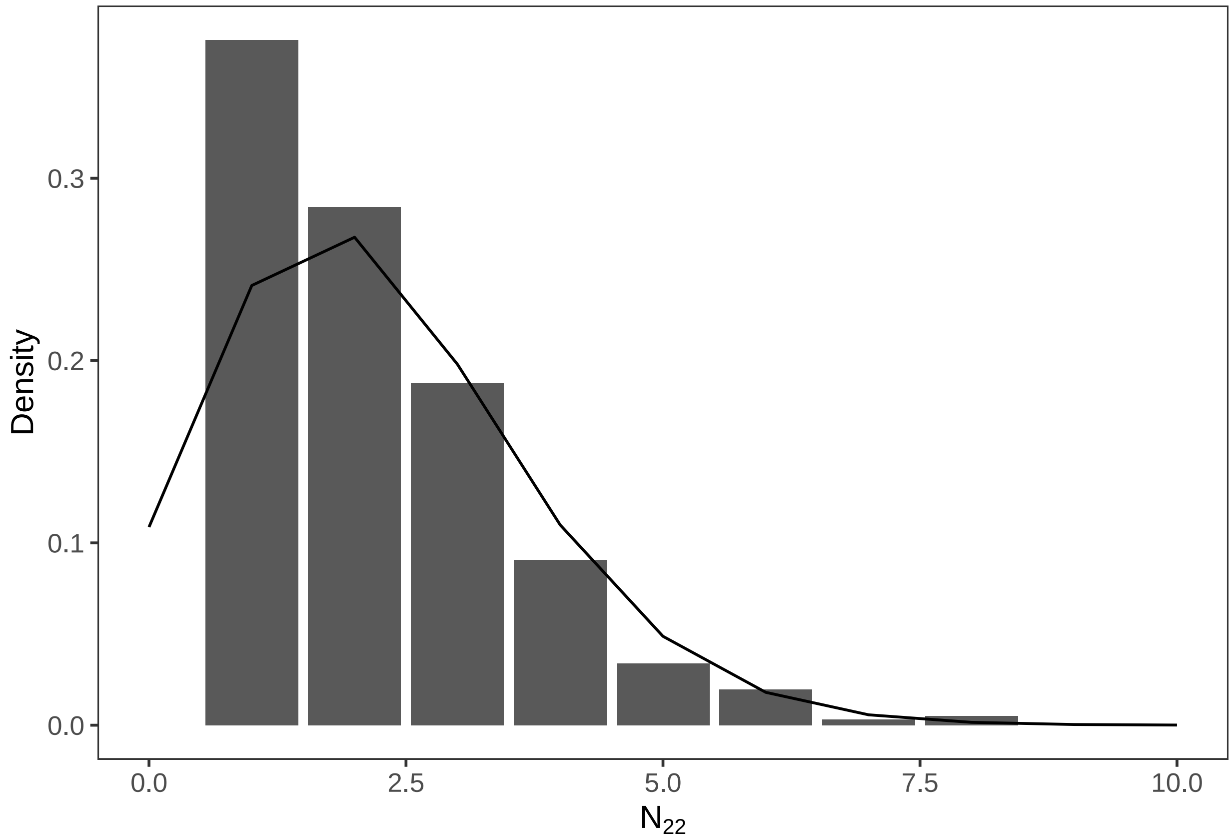
$$N_{ij} \sim \text{Po}(\lambda_{ij})$$


0.125
0.100
0.075
0.050
0.025
0.000
Density
0
10
20
30
N_{11}


0.25
0.20
0.15
0.10
0.05
0.00
Density
0
5
10
15
N12

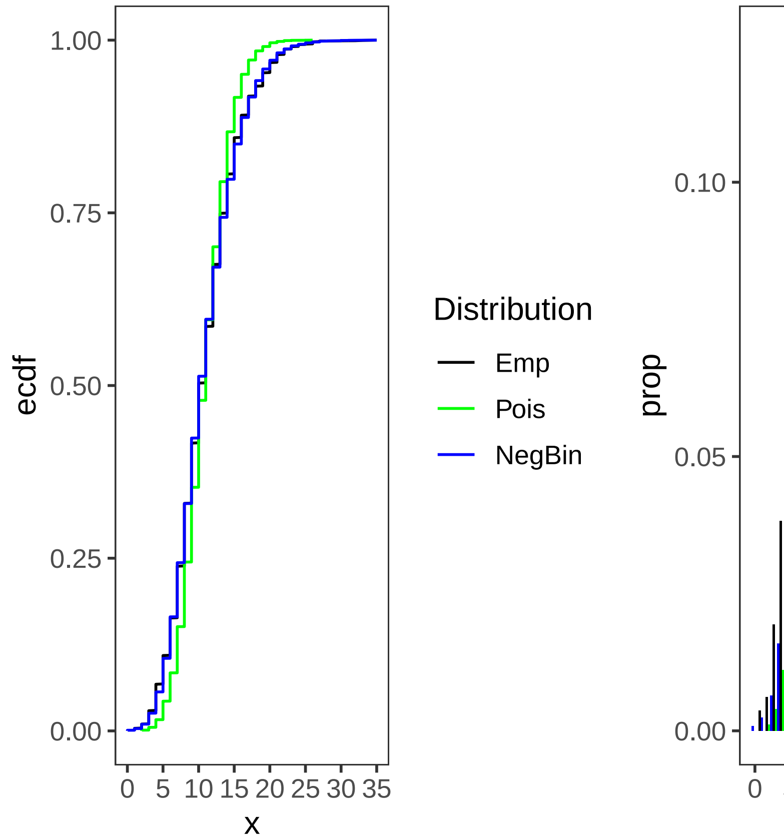


0.15
0.10
0.05
0.00
Density
0
5
10
15
20
N_{21}

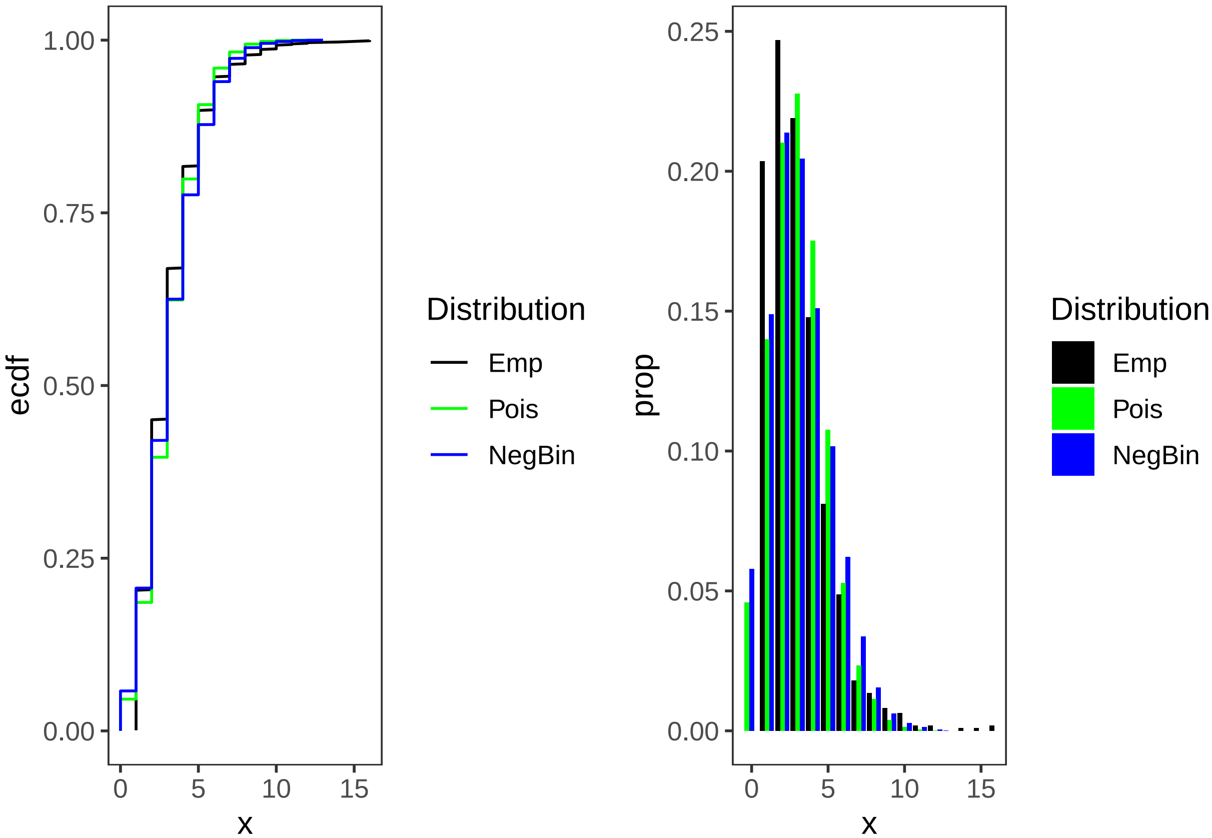


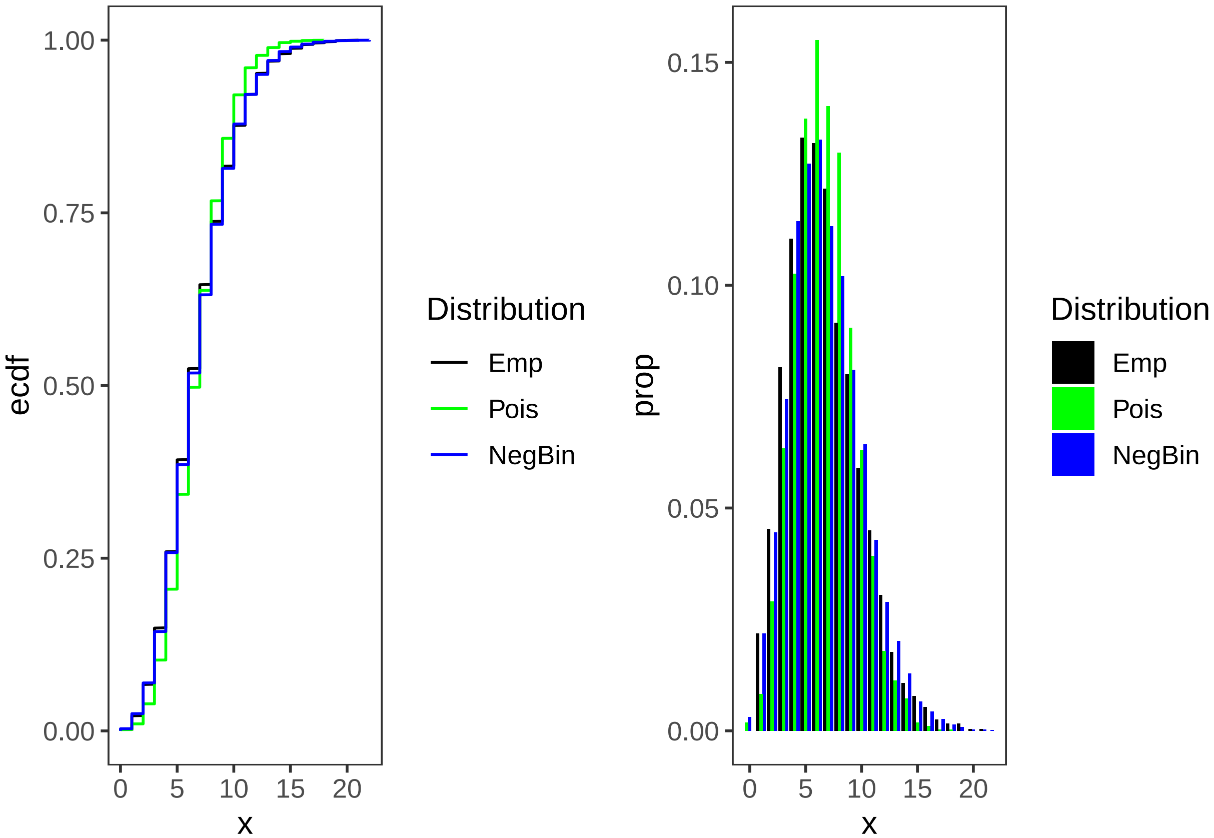
The Poisson variables which have been fit for the different seasons and insurance branches have had their parameter λ_{ij} estimated through maximum-likelihood methods using the data from the 10 previous years. We note that we seem to have overdispersion for N_{11} and N_{12}, meaning that the variance is not truly equal to the expectation which is the case for a Poisson variable. For N_{21} and N_{22} date does however seem to indicate that we have underdispersion, meaning that the variance is lower than the expectation.

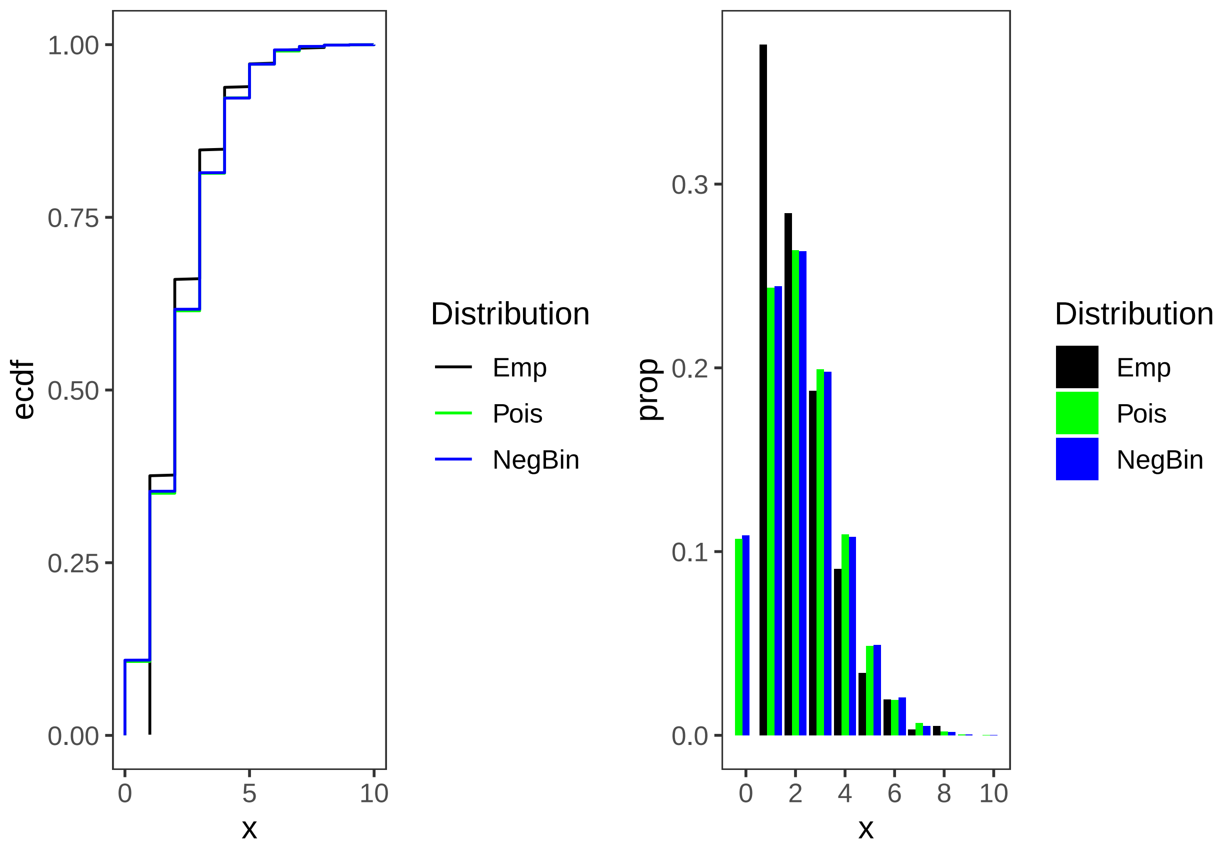
To solve this we fit negative binomial distributions to N_{ij} as this distribution does not have the property of equal



expectation and variance as the Poisson distribution does.

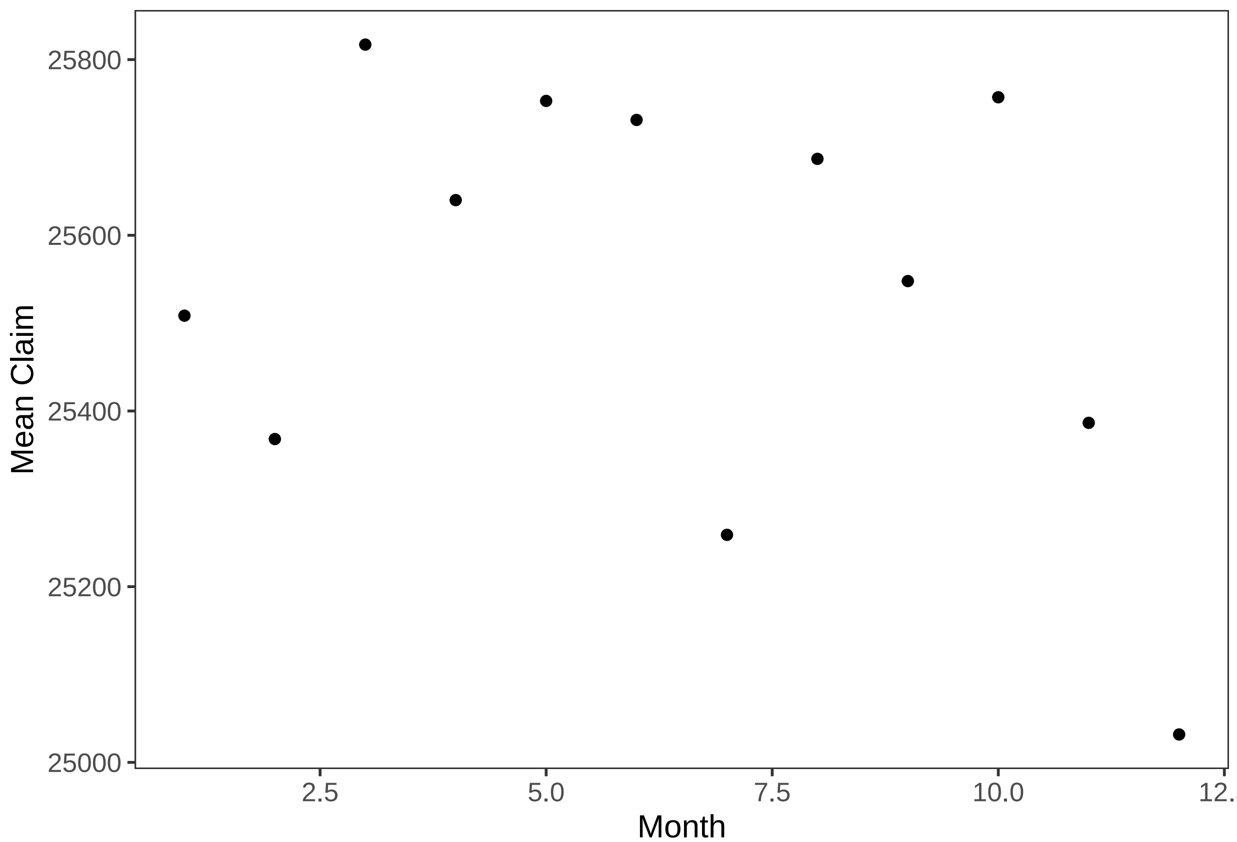

ecdf
1.00
0.75
0.50
0.25
0.00
0
5
10
15
x
Distribution
Emp
Pois
NegBin
prop
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
x
Distribution
Emp
Pois
NegBin

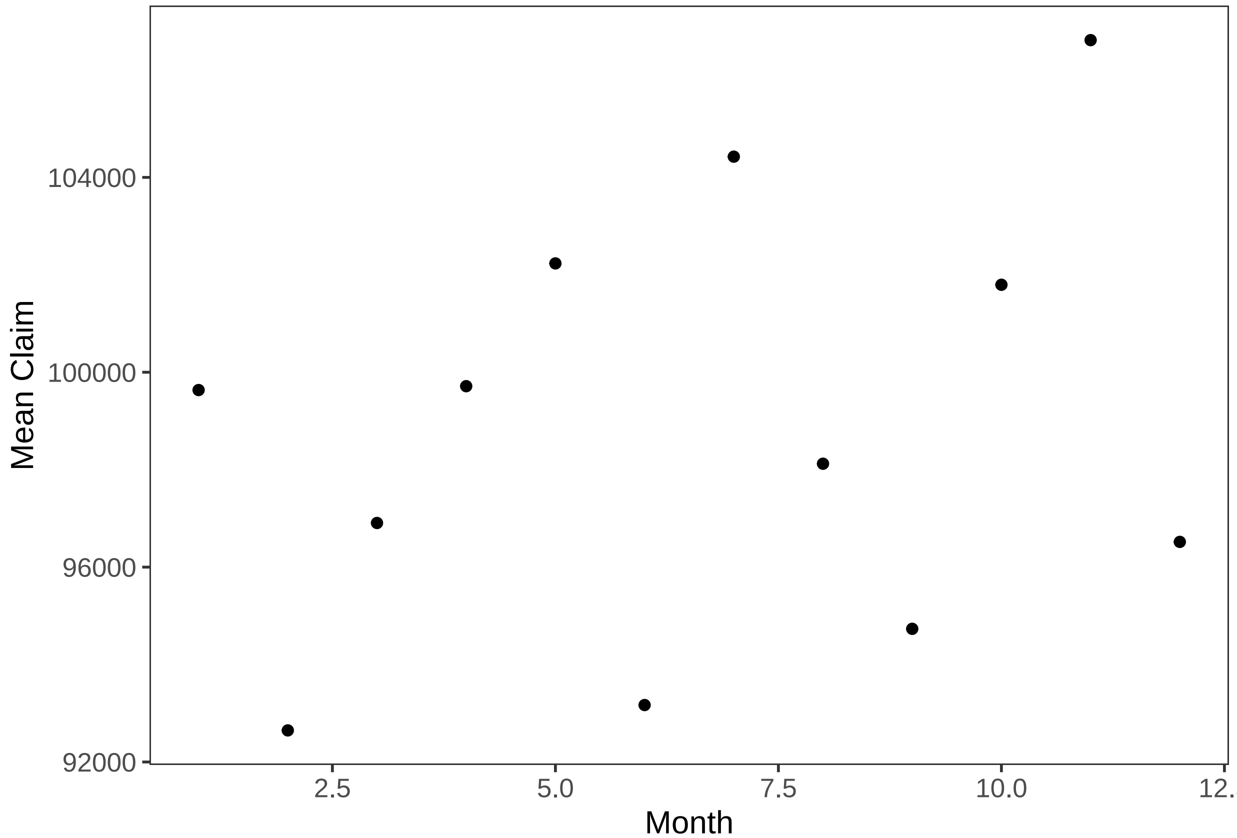

ecdf
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
x
Distribution
Emp
Pois
NegBin
prop
0.15
0.10
0.05
0.00
0
5
10
15
20
x
Distribution
Emp
Pois
NegBin



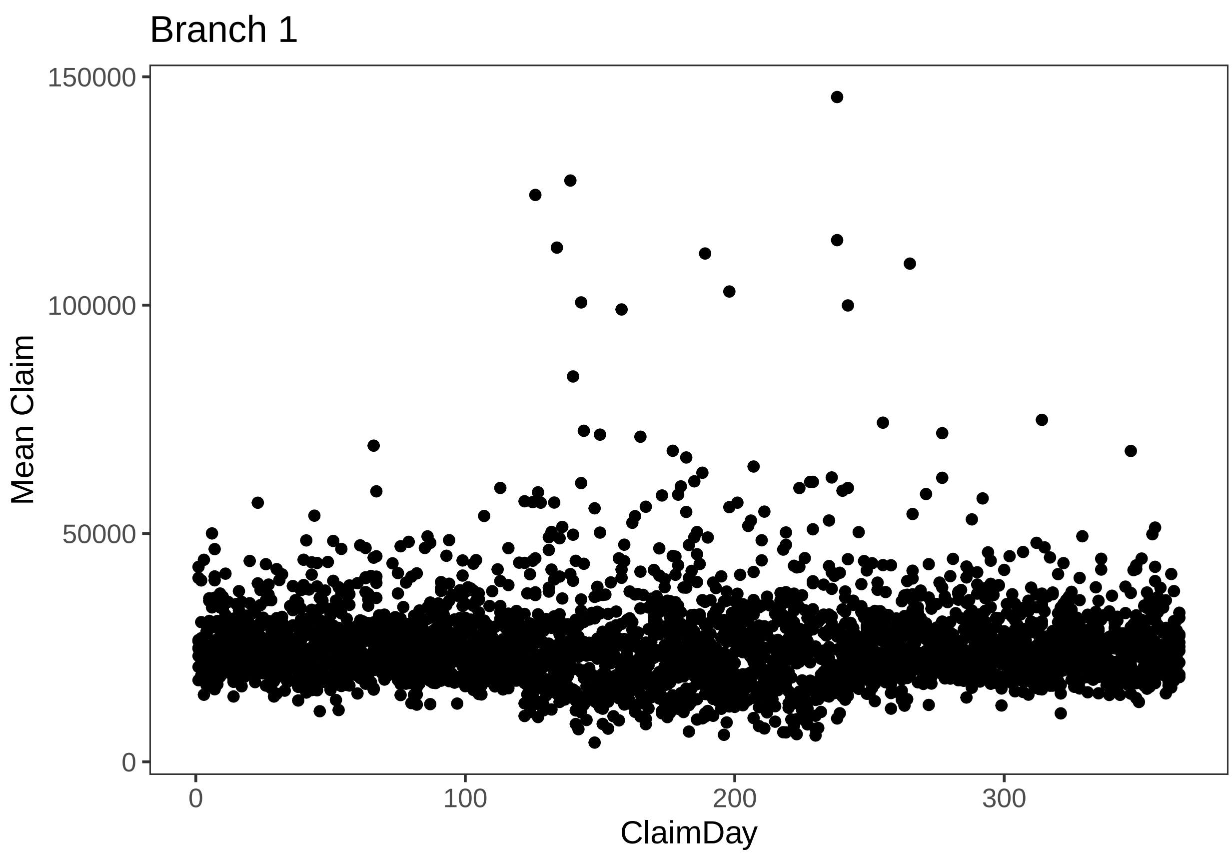
Exercise 2

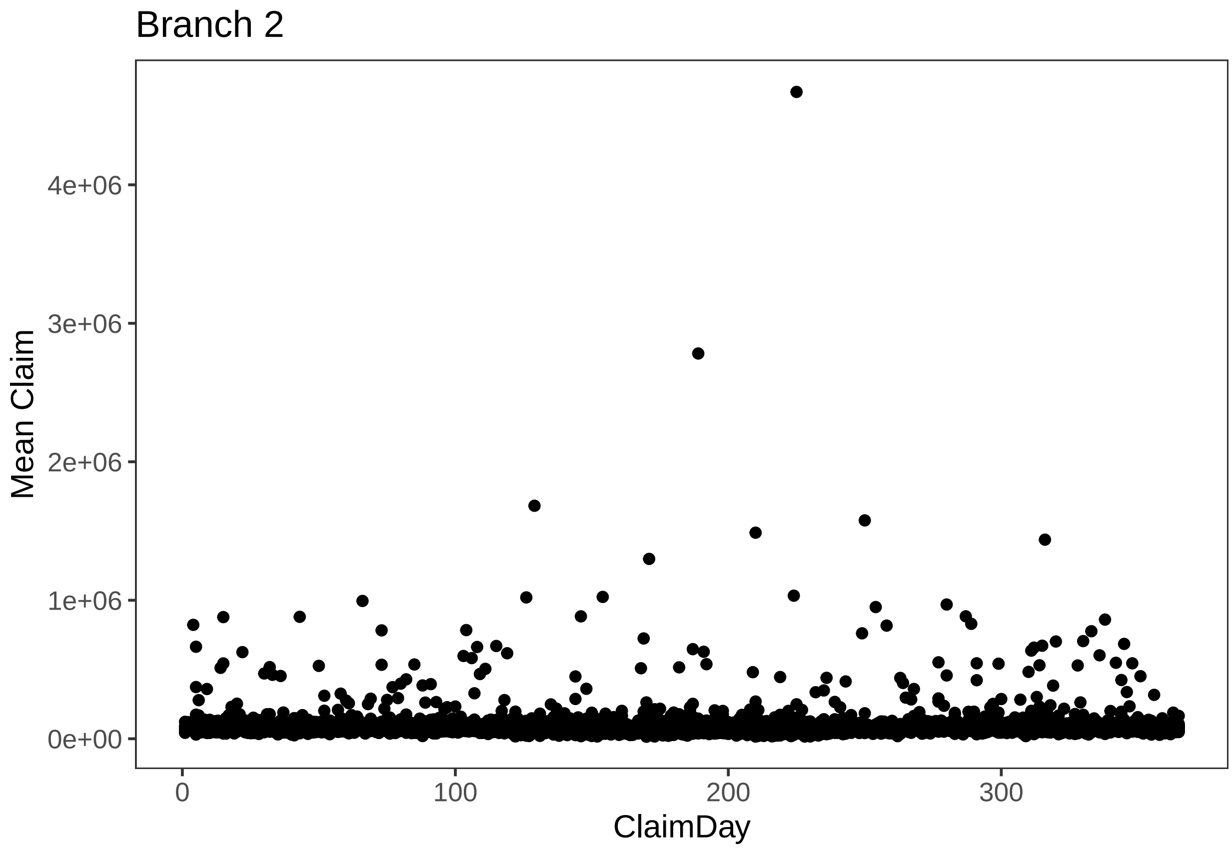
It is not only of interset to observe the number of claims over time but to also examine the sizes of these claims.


Mean Claim
25800
25600
25400
25200
25000
2.5
5.0
7.5
10.0
12.5
Month

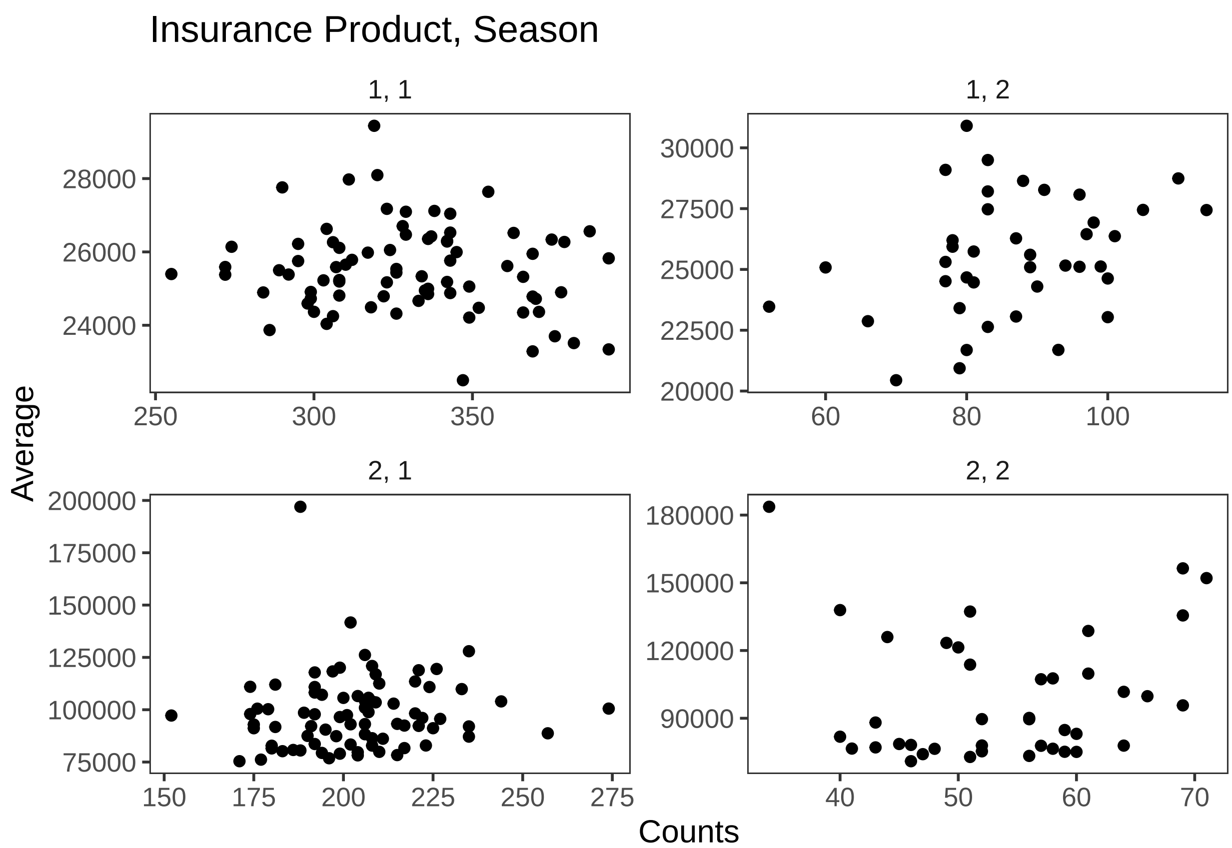


It seems as if the avergae claim costs are time independent, atleast when aggregated on a monthly basis, from the two previous plots. We can also observe the mean of the claim sizes for every day in order to further strengthen the assumption of time independence.


Branch 1
150000
100000
50000
0
Mean Claim
0
100
200
300
ClaimDay

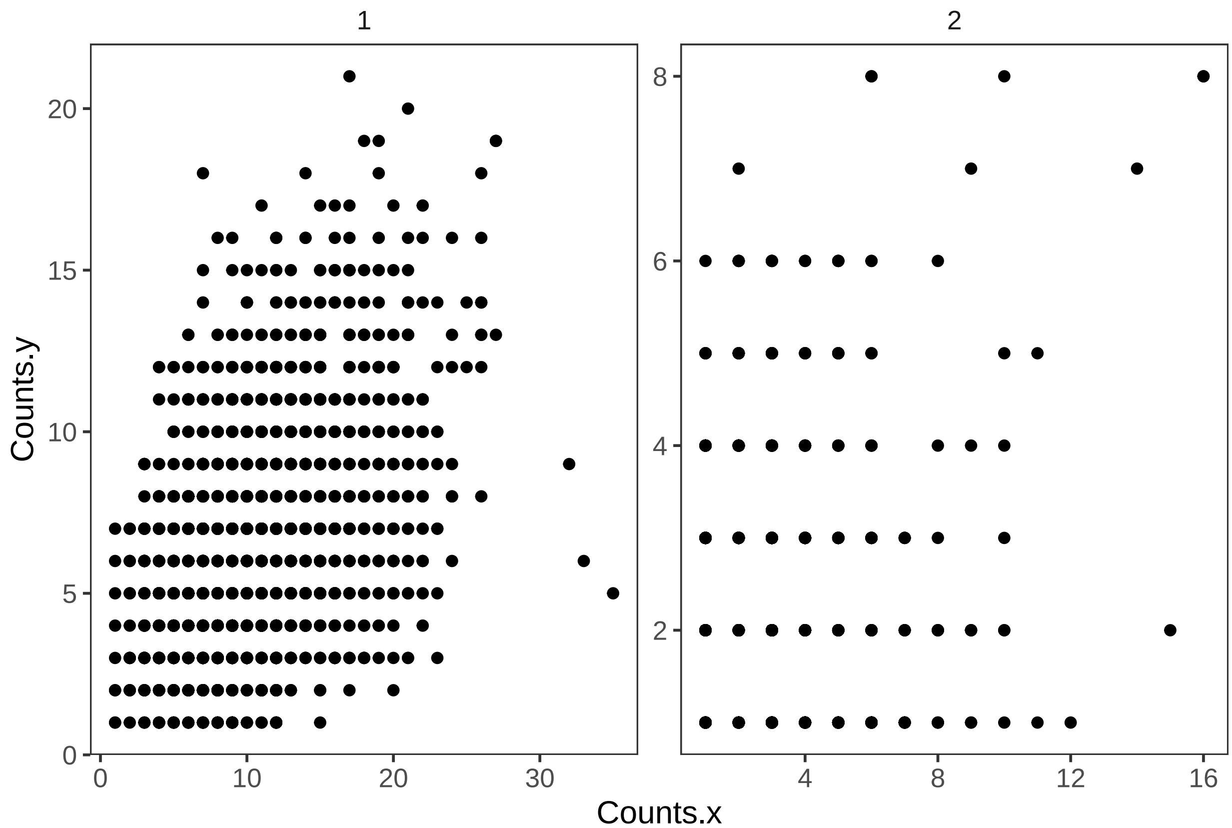


We note from the figures that the average claim sizes seem to be time independet. An interesting factor is however that all extremely deviating claim costs seem to occur during the summer (i.e period 2). In conclussion it seems as if the claim costs are independent of time. Another thing that we want to examine is whether or not we have independence between the average claim costs and the number of claims.



We note that we do not seem to have any systematic correlation for any of the combinations of claims and claim costs, meaning that we can model the number of claims and the claim costs independently.

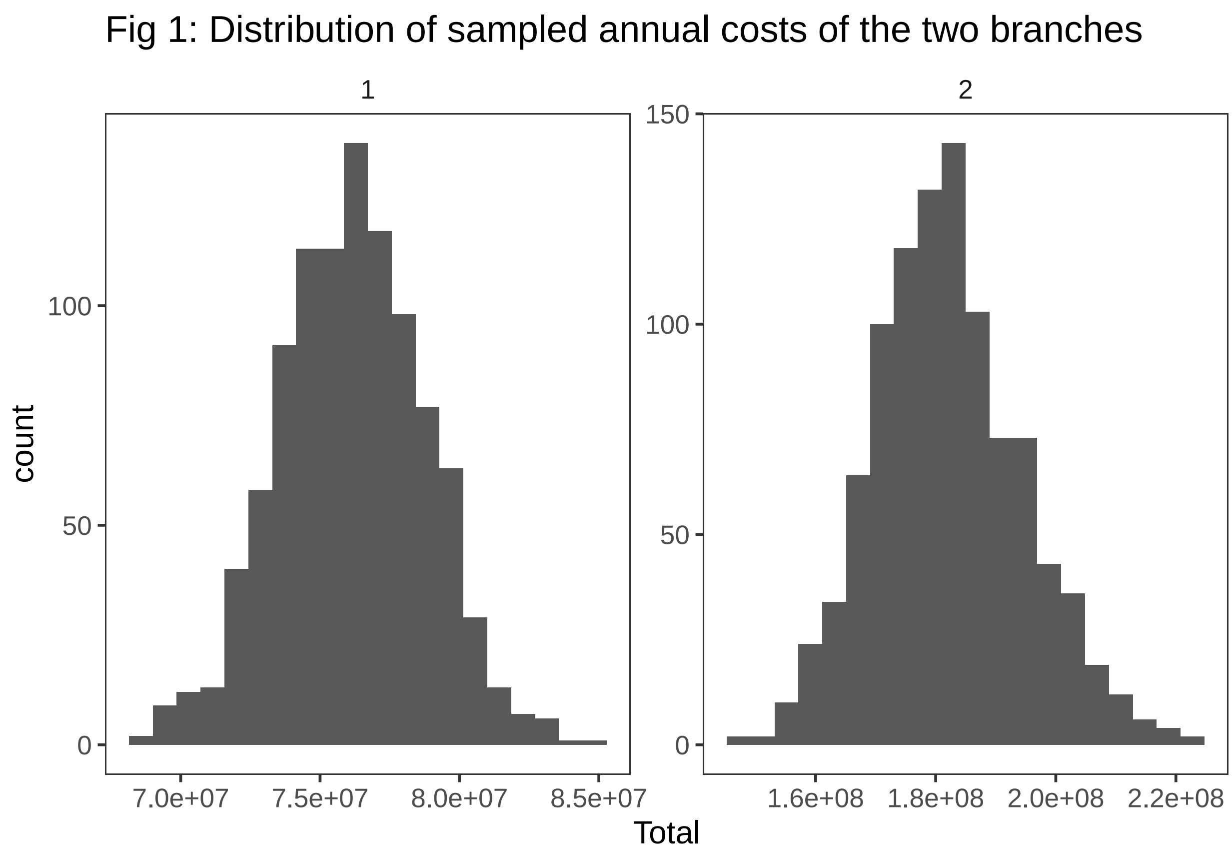
Exercise 3



We note from the figures above that there seems to be some kind of positive correlation between the two insurance products, meaning that we cannot model the claims for each product separatley, but that we instead have to model the jointly. We can however still model the claim costs of the two branches independently as we saw in the previous exercise.

Exercise 4

As we mentioned previously the number of claims for the two branches are not independent, meaning that the have to be sampled from a bivariate distribution rather than two univariate distributions. This can be done by either deriving the bivariate distribution analytically and then sampling from it or, as we will do, by using a bootstrap sampler. We begin by sampling the number of claims for each month of the following year from the data of the correct season.

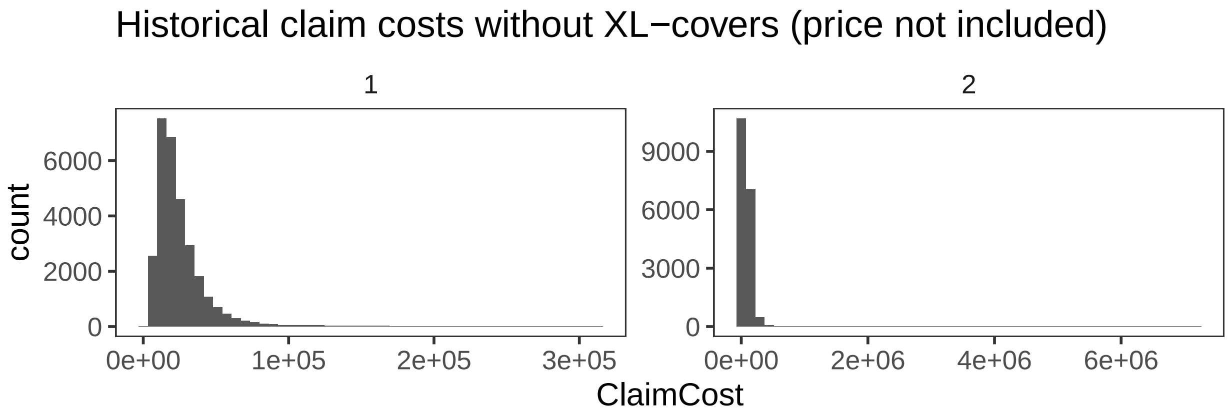


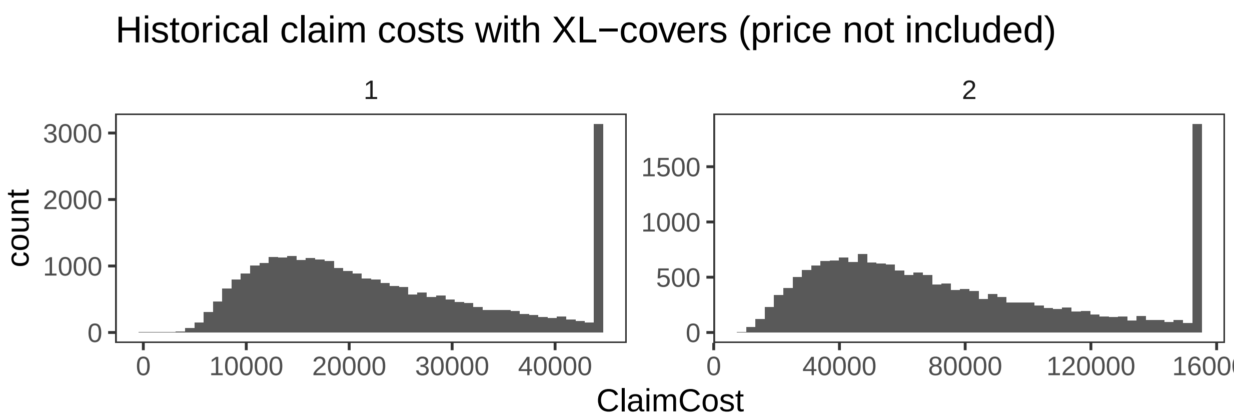
Exercise 5

We now want to implement two separate XL-covers. The first cover caps losses for the 10% worst claims. For our two branches these cut-off M_1 and M_2 are the following

Table 1: Cut-offs

Type	M
1	44562
2	154011



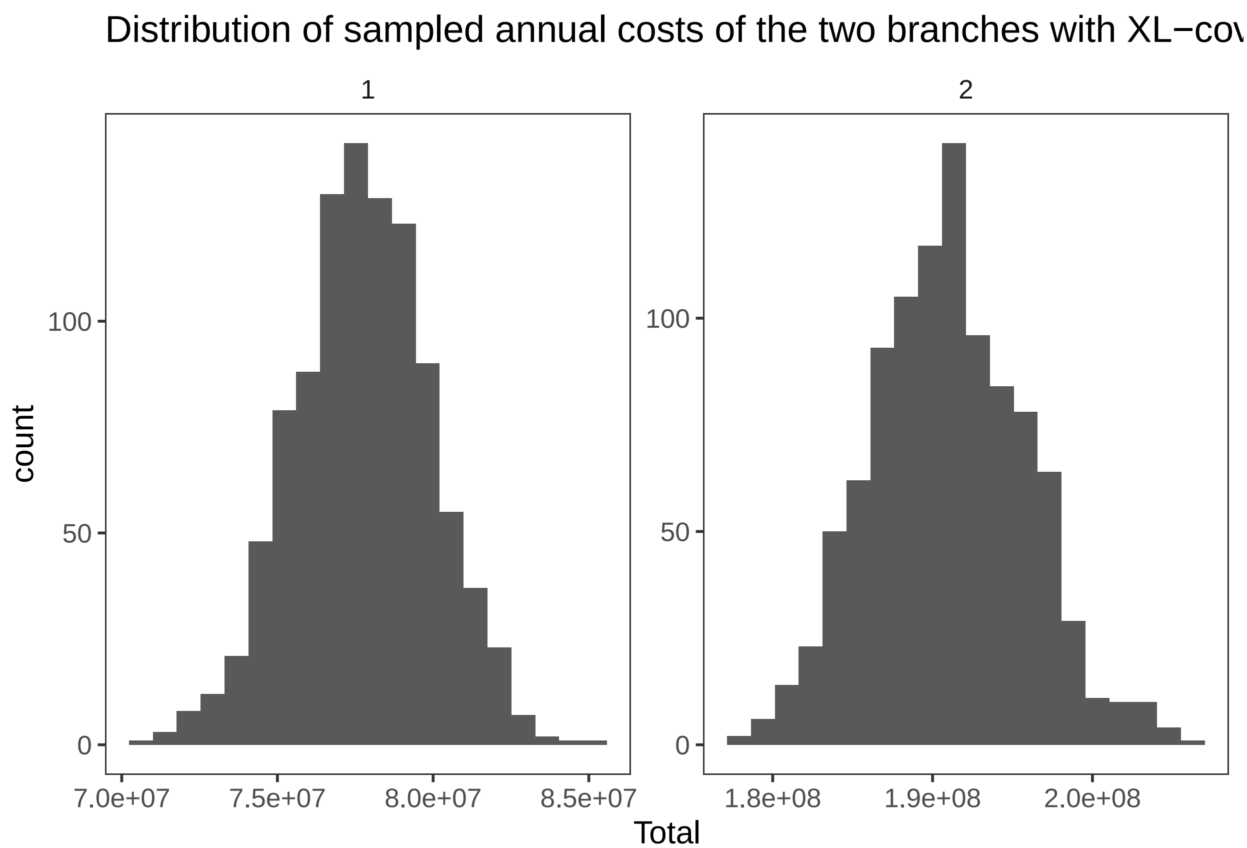


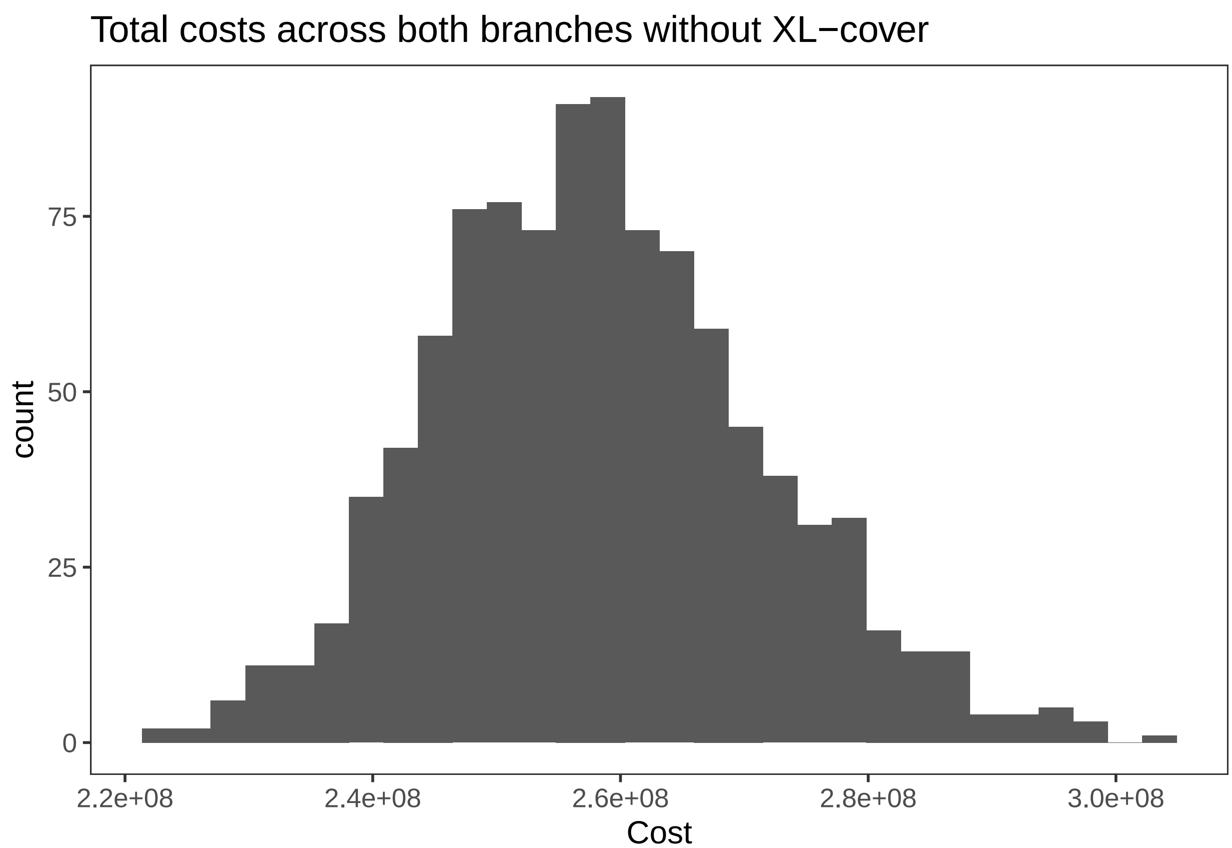
From the figure above we see the distribution of claim costs for the two branches with, and wihtout, XL-cover. Of course these covers come at a price which we can estimate from histrocial data.

Table 2: Prices for the two XL-covers

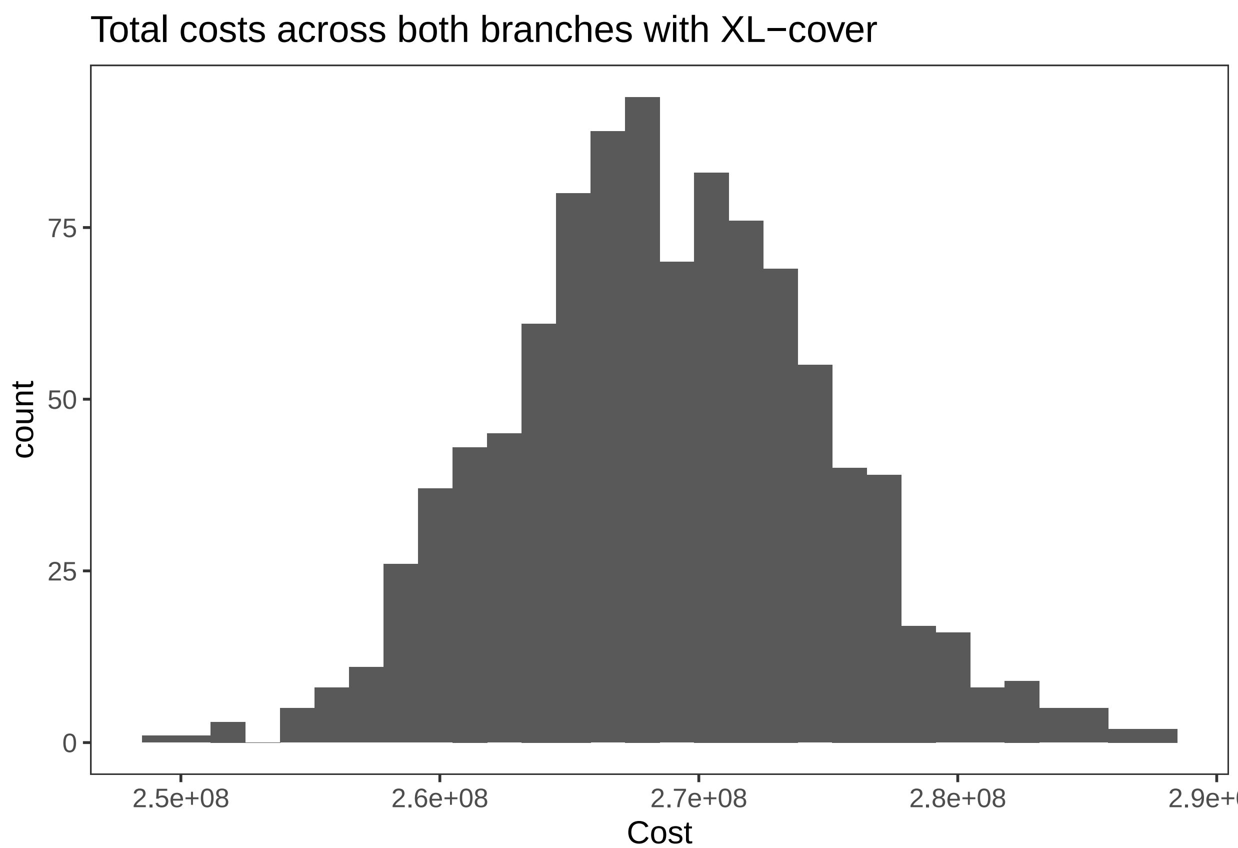
Type	Price
1	9473202
2	53191862

We then want to examine how the purchase of these XL-covers would impact our total losses for the forthcoming year. We saw a simulated distirbution of the losses without the XL-covers in figure 1. If we implement the covers we get the following simulated distributions.


Distribution of sampled annual costs of the two branches with XL−cov
1
2
count
100
50
0
100
50
0
7.0e+07
7.5e+07
8.0e+07
8.5e+07
1.8e+08
1.9e+08
2.0e+08
Total



Total costs across both branches without XL−cover
count
0
25
50
75
2.2e+08
2.4e+08
2.6e+08
2.8e+08
3.0e+08
Cost



We see that the distribution is lighter-tailed which of course is to be expected after the purchase of an XL-cover.

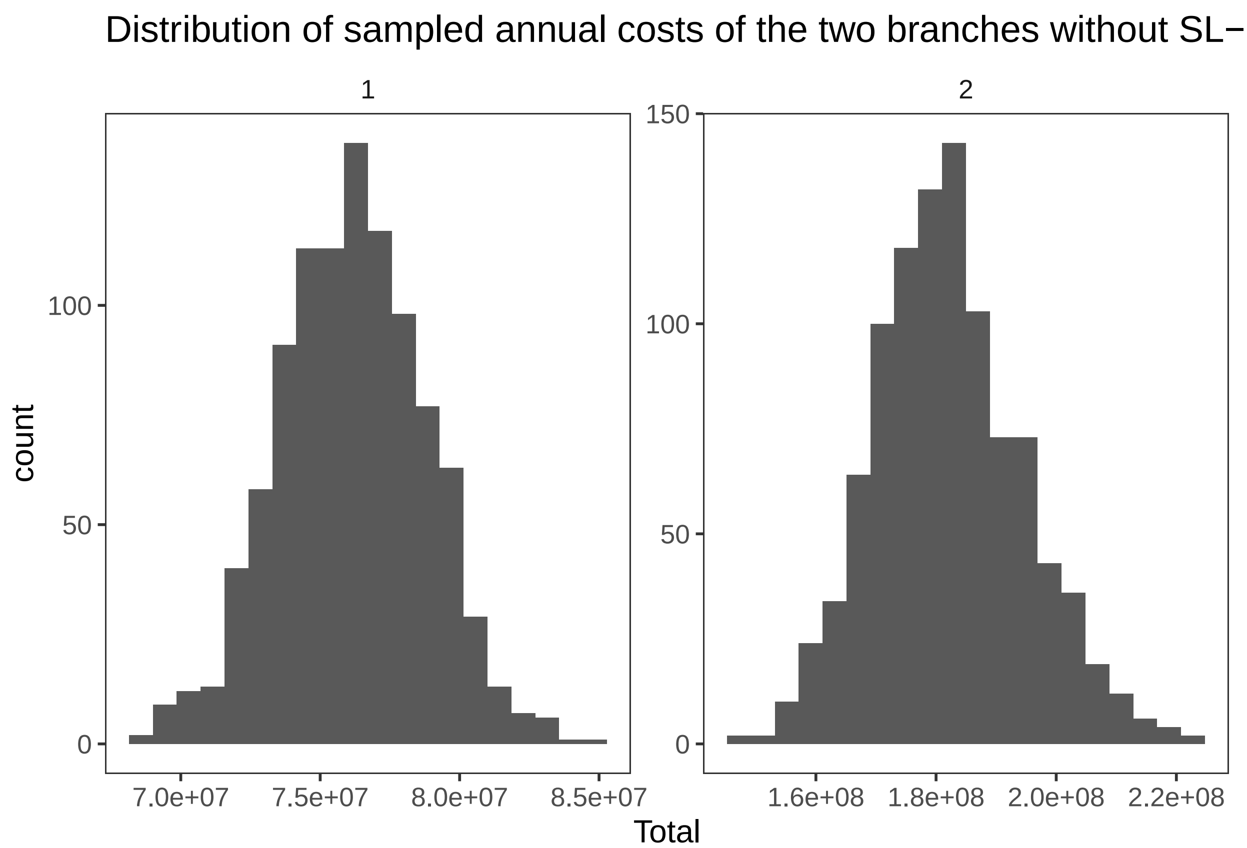
Exercise 6

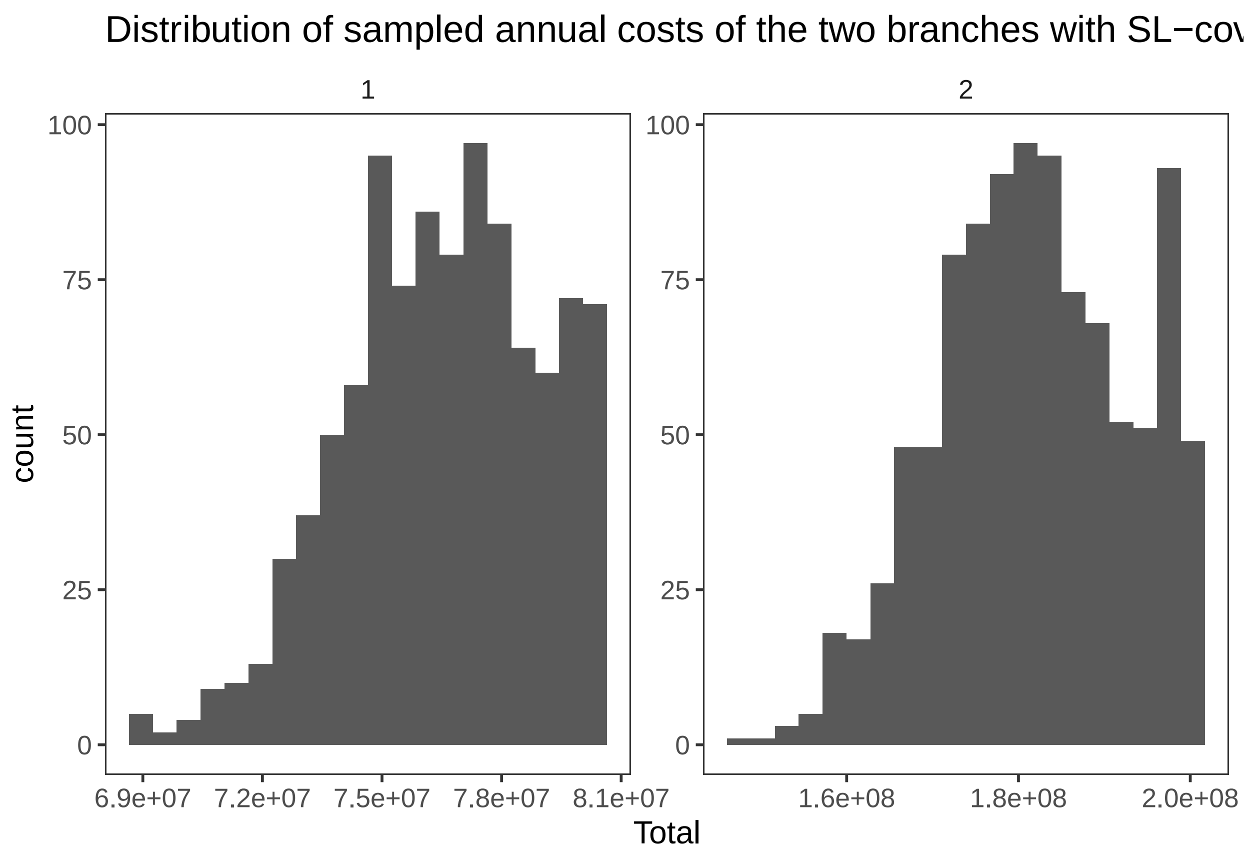
We now want to implement an SL-cover instead of an XL-cover for both insurance branches. Like the XL-covers the SL-covers insure against the 10% worst annual costs at a price of 120% of the expected cost.

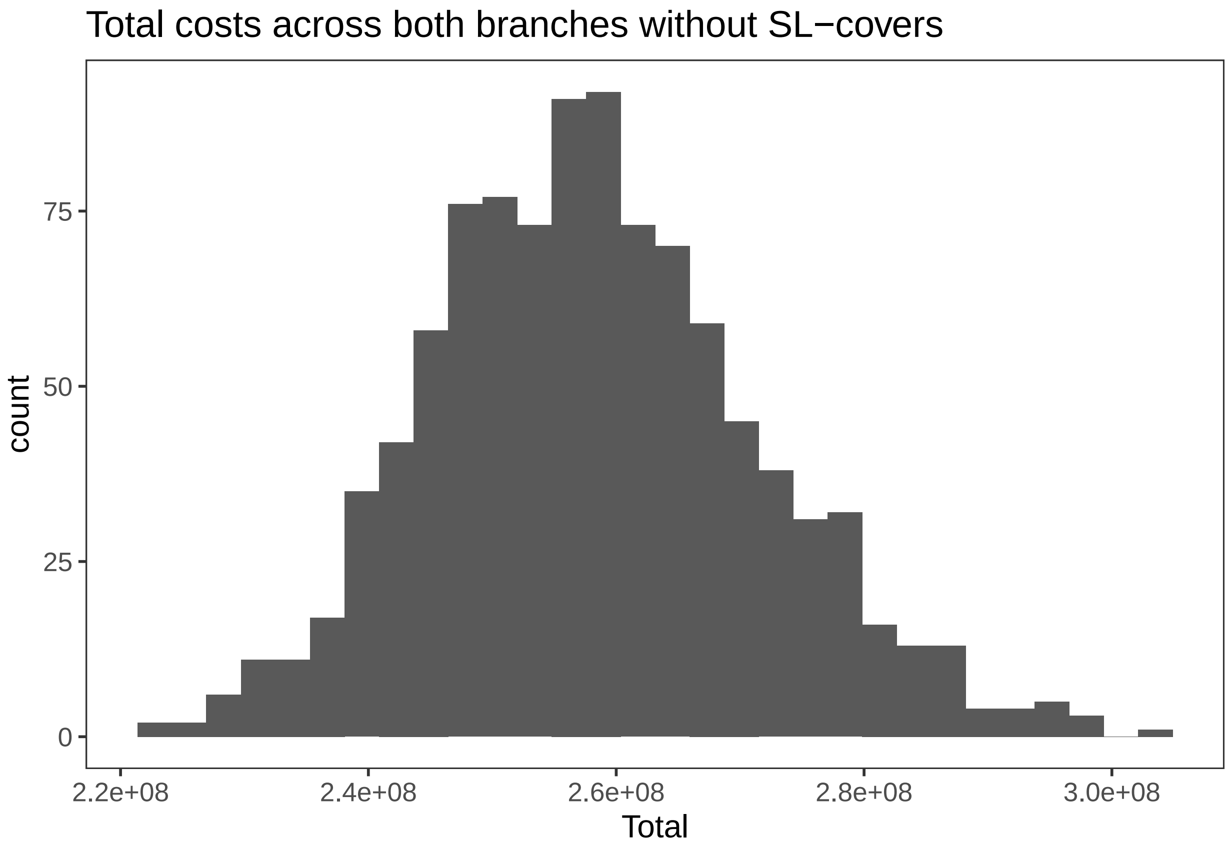
The prices of these two SL-covers do of course come at a cost.

Table 3: Prices for the two SL-covers

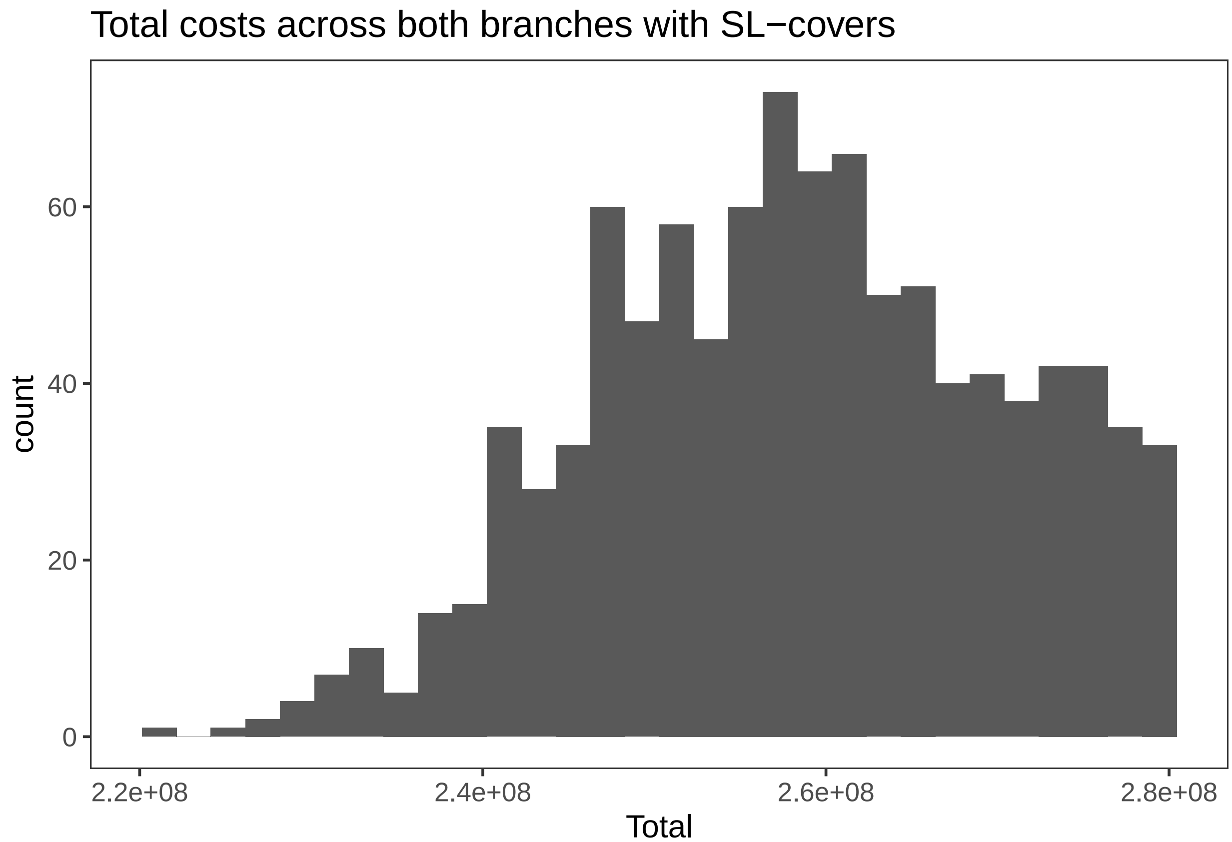
Type	Price
1	136864.5
2	839111.3


Distribution of sampled annual costs of the two branches without SL−
1
2
count
150
100
50
0
7.0e+07
7.5e+07
8.0e+07
8.5e+07
1.6e+08
1.8e+08
2.0e+08
2.2e+08
Total


Distribution of sampled annual costs of the two branches with SL−cov
1
2
100
75
50
25
0
count
6.9e+07
7.2e+07
7.5e+07
7.8e+07
8.1e+07
1.6e+08
1.8e+08
2.0e+08
Total



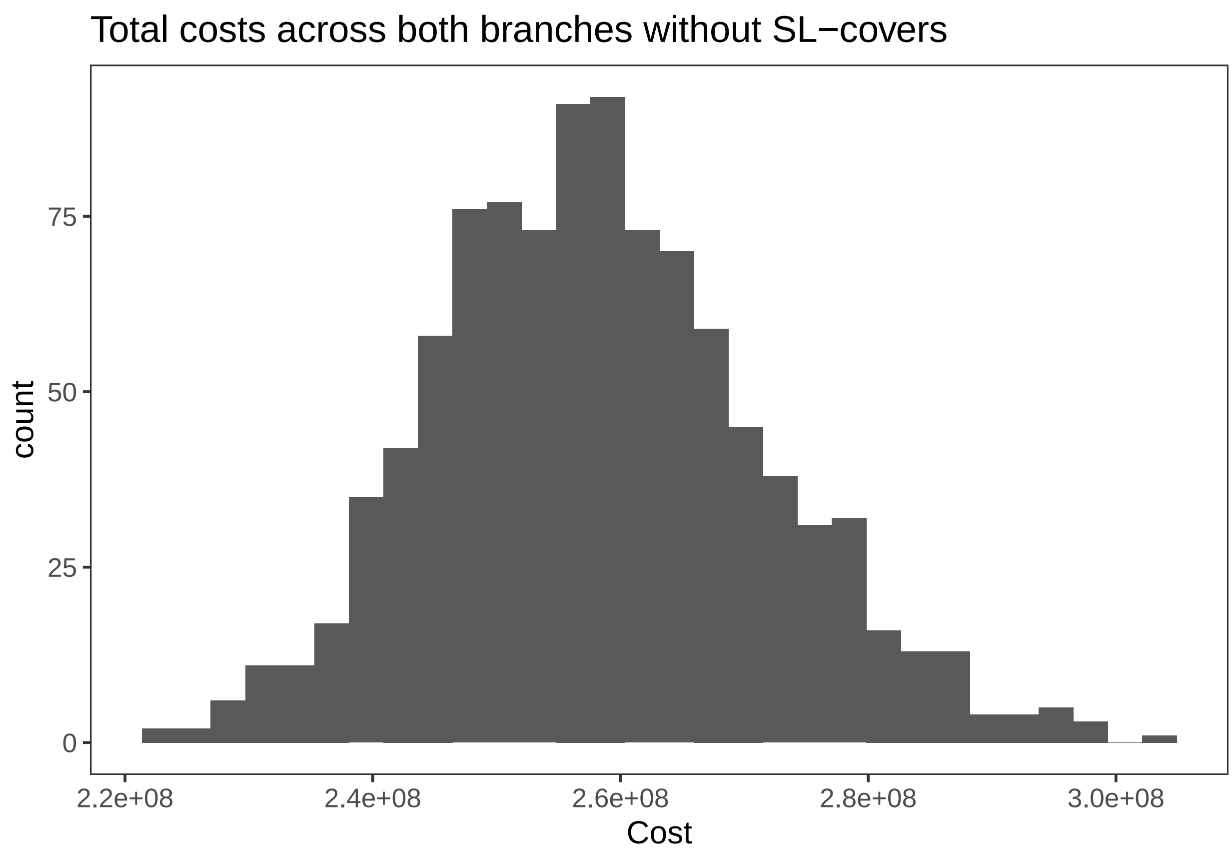
Total costs across both branches without SL−covers
count
75
50
25
0
2.2e+08
2.4e+08
2.6e+08
2.8e+08
3.0e+08
Total

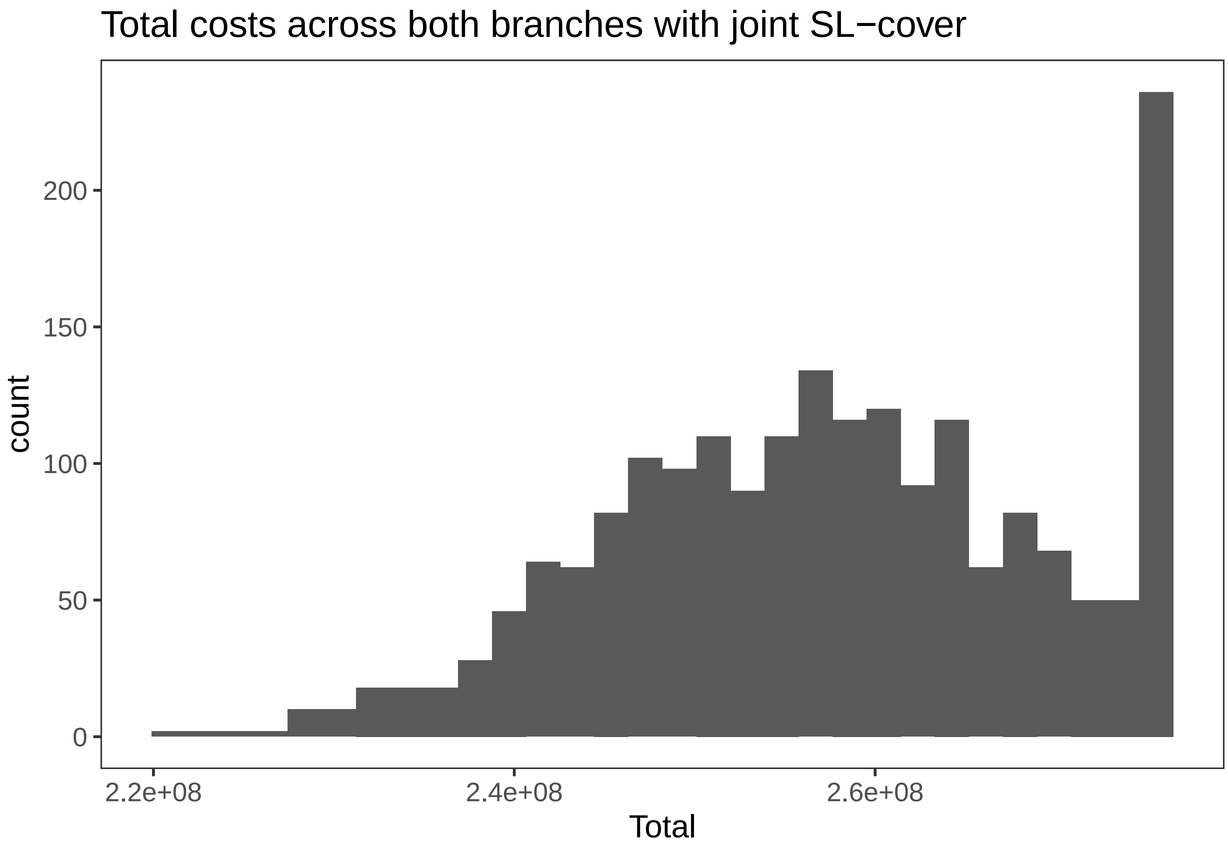


Exercise 7

Lastly we want to implement an SL-cover which insured against 10% of the worst total annual costs, i.e. summed over both branches. This cover will be priced and insure like the individual SL-covers in Exercise 6.

The cost of this cover, based on historical data, is 8.3657639×10^5


Total costs across both branches without SL−covers
count
75
50
25
0
2.2e+08
2.4e+08
2.6e+08
2.8e+08
3.0e+08
Cost



We see from the two histograms that the total SL-cover implies a higher expected cost but also limits the total annual cost to $M = 2.7615045 \times 10^8$. As such a more risk averse insurance provider might opt for the SL-cover to completly exclude the possiblity of large deviations in the annual costs whilst a more risk-taking insurance provide might avoid the SL-cover to reduce the expected annual cost.