MT7027: Project 1

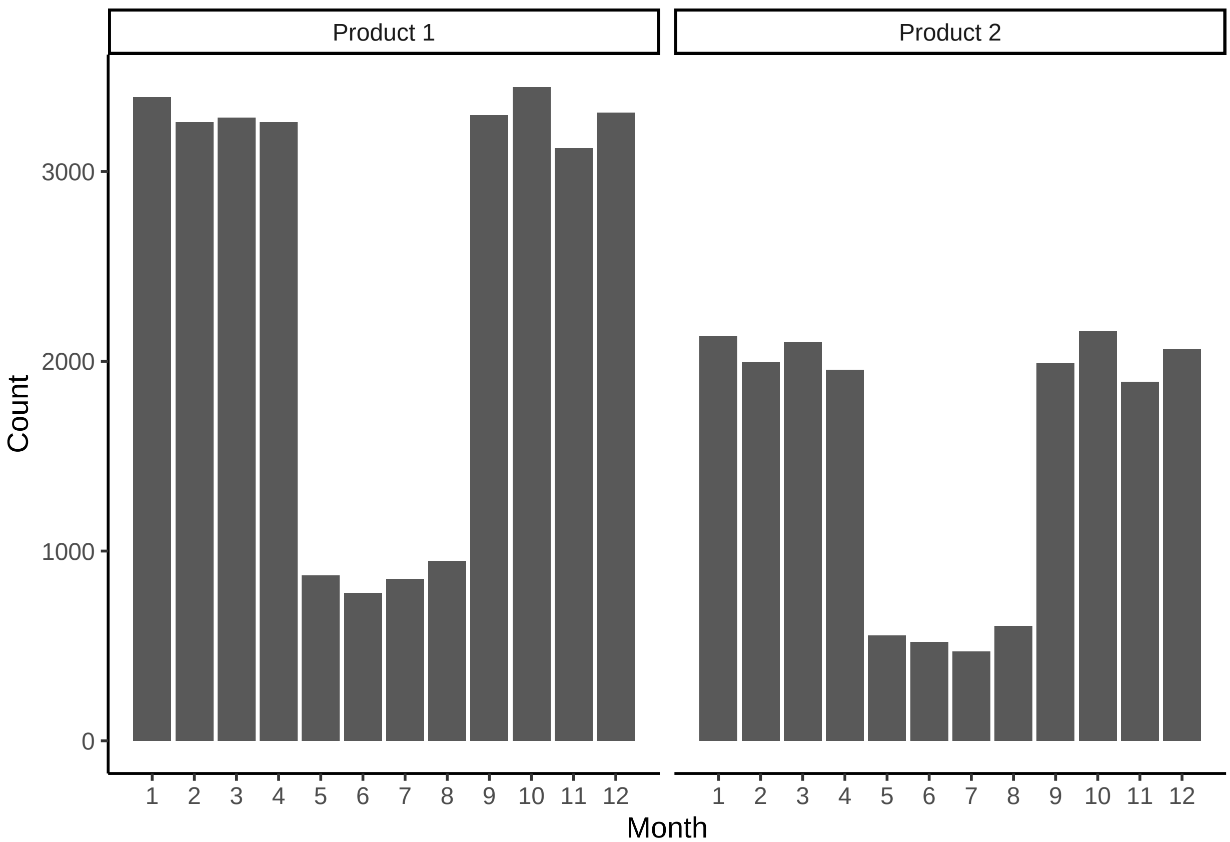
Anton Stråhle, Jan Alexandersson & Max Sjödin

Introduction

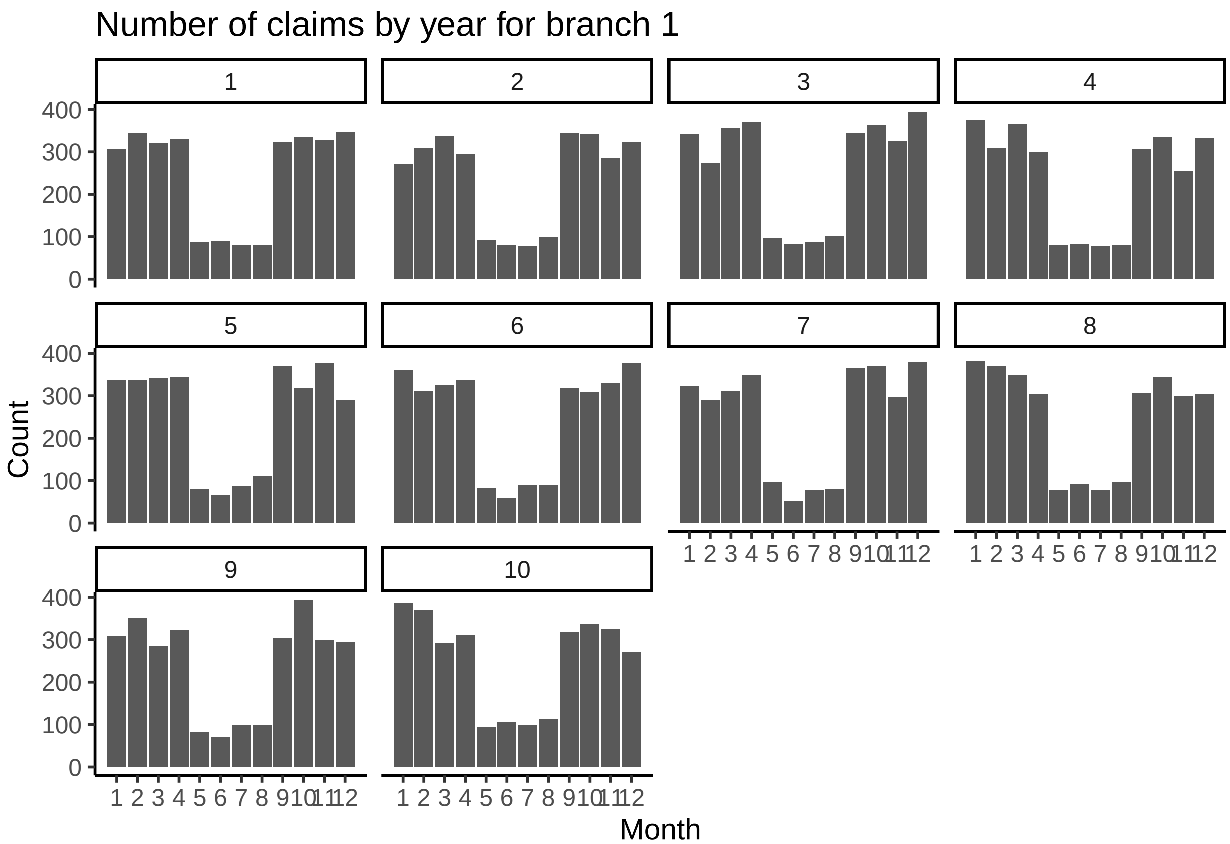
In this project we are dealing with data concerning two different insurance branches collected over 10 years. We do not know anything about the sizes of the two insurance portfolios except the fact that their size has not changed over the decade which the data spans. Furthermore, the insurance products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

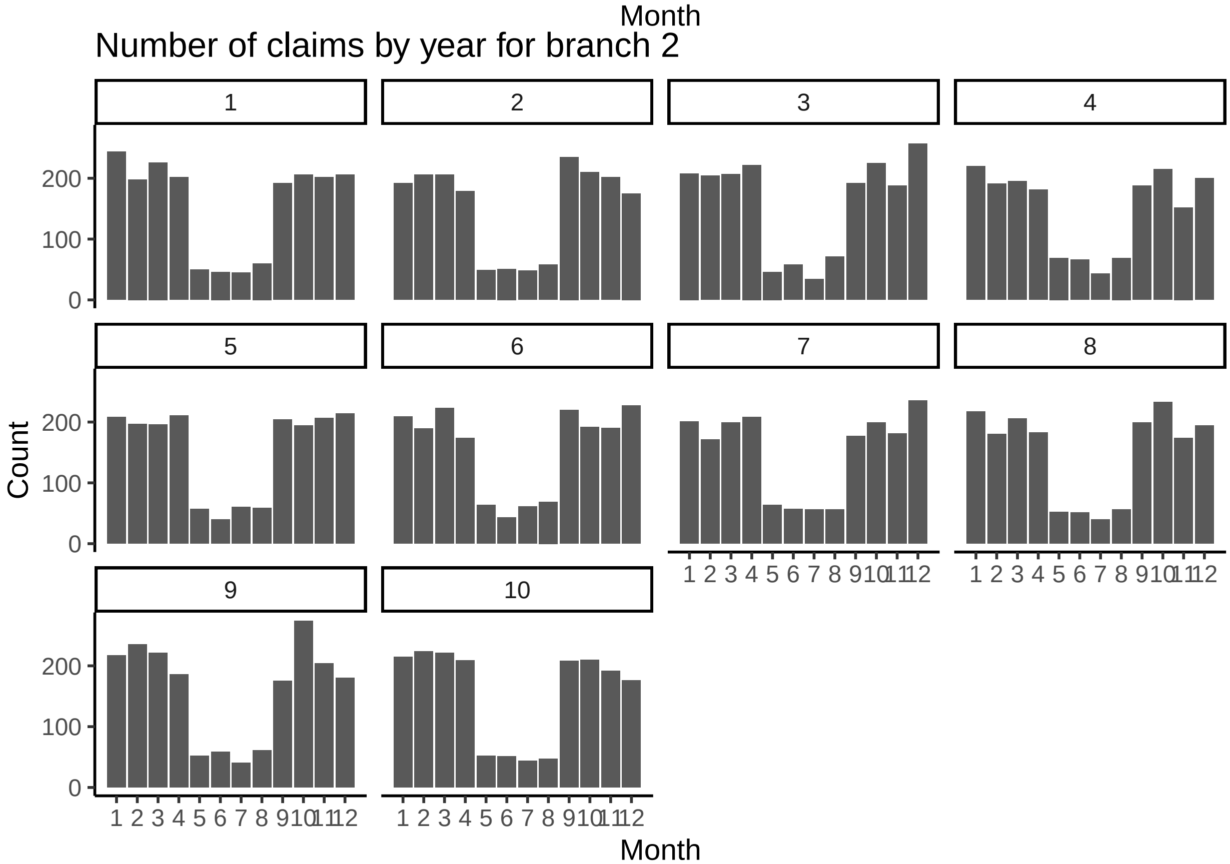
Exercise 1

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.





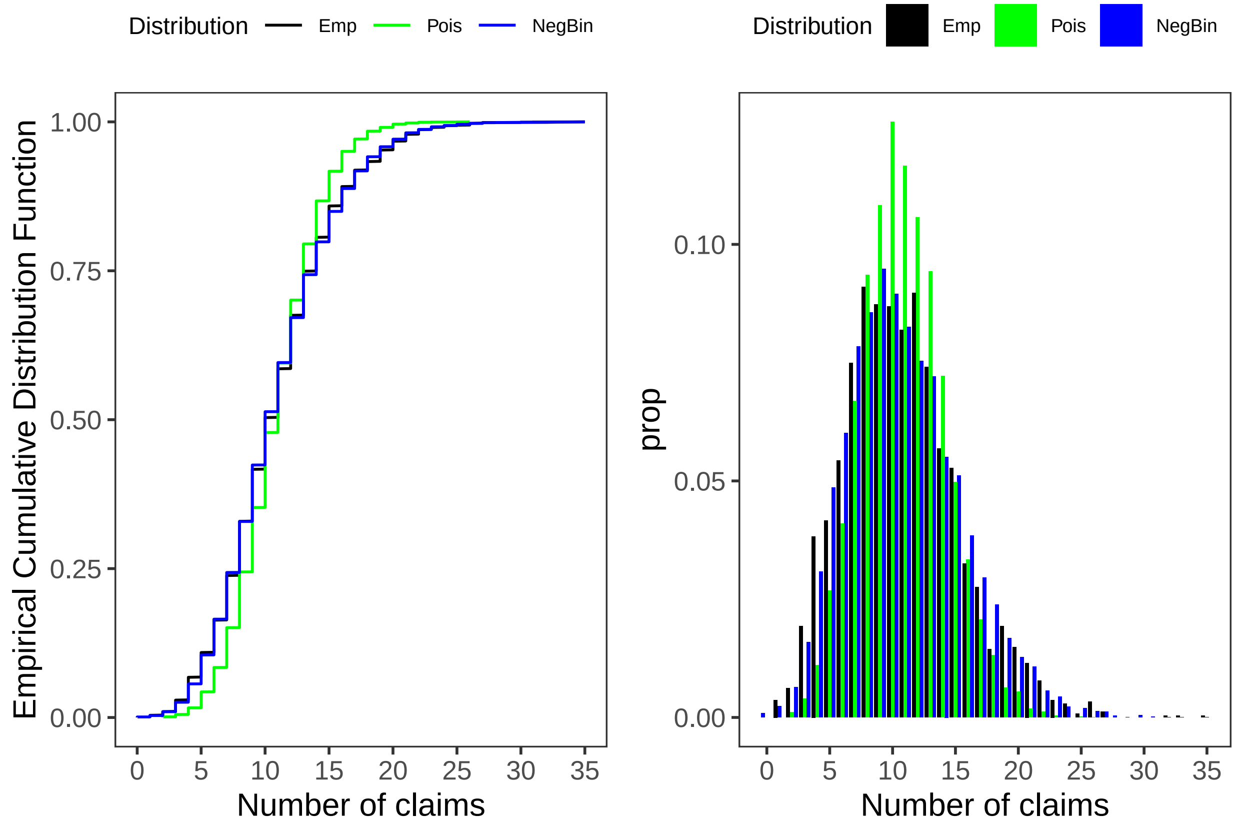
We note from the previous plots that there does not seem to be any major difference in the number of claims between the years.

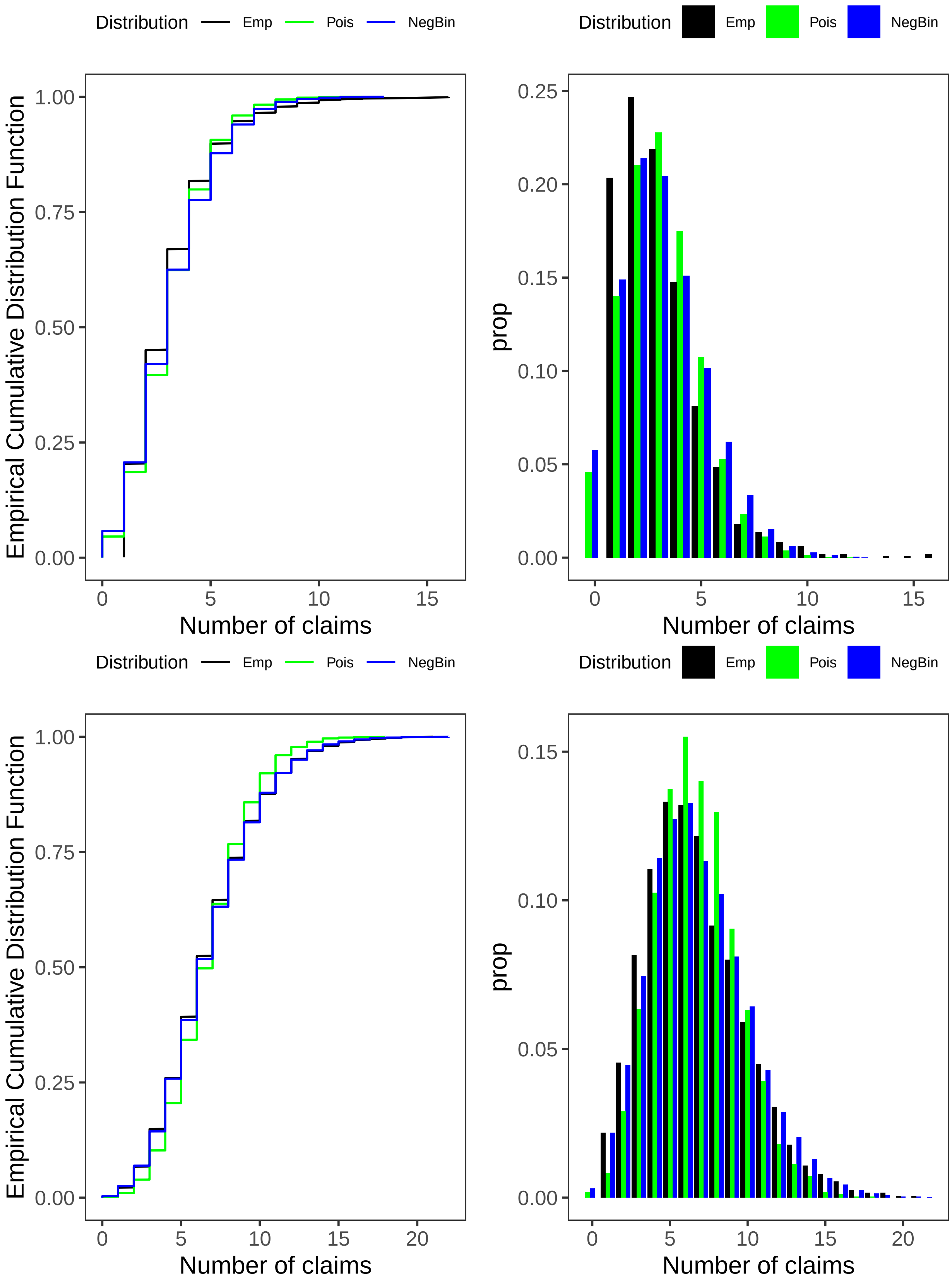
We now wish to fit homogeneous Poisson distributions N_ij (where i represents the insurance type and j represents the season) to the months of each season and each insurance product.

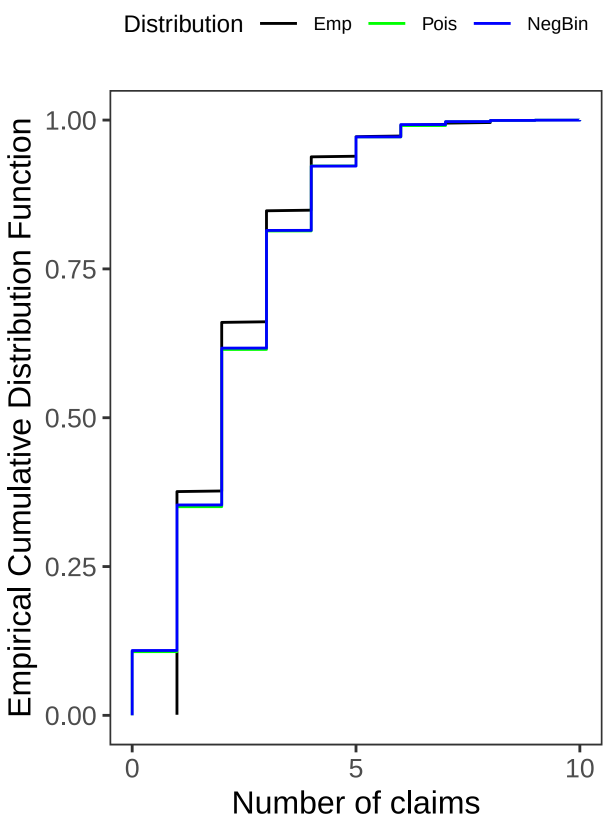
$$N_{ij} \sim \mathrm{Po}(\lambda_{ij})$$

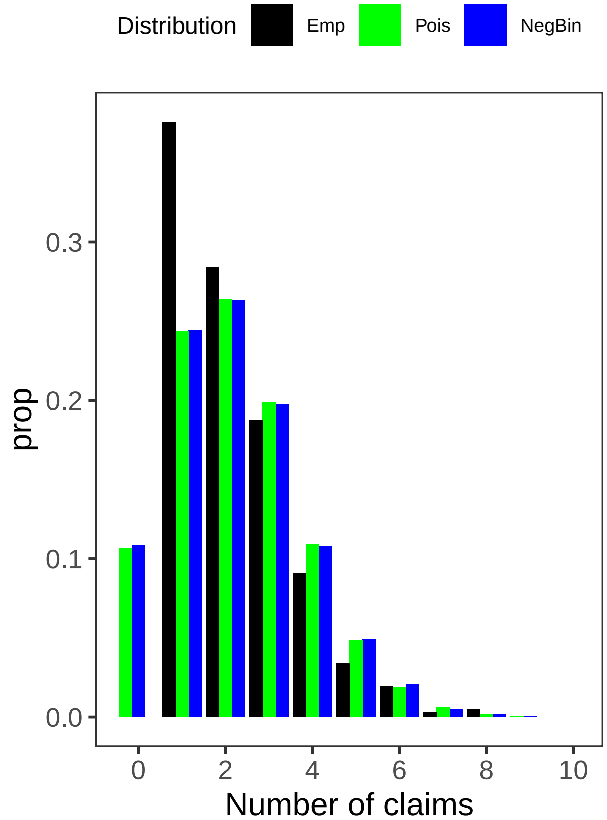
The Poisson variables which have been fit for the different seasons and insurance branches have had their parameter λ_{ij} estimated through maximum-likelihood methods using the data from the 10 previous years. We note that we seem to have overdispersion for N_{11} and N_{12}, meaning that the variance 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.



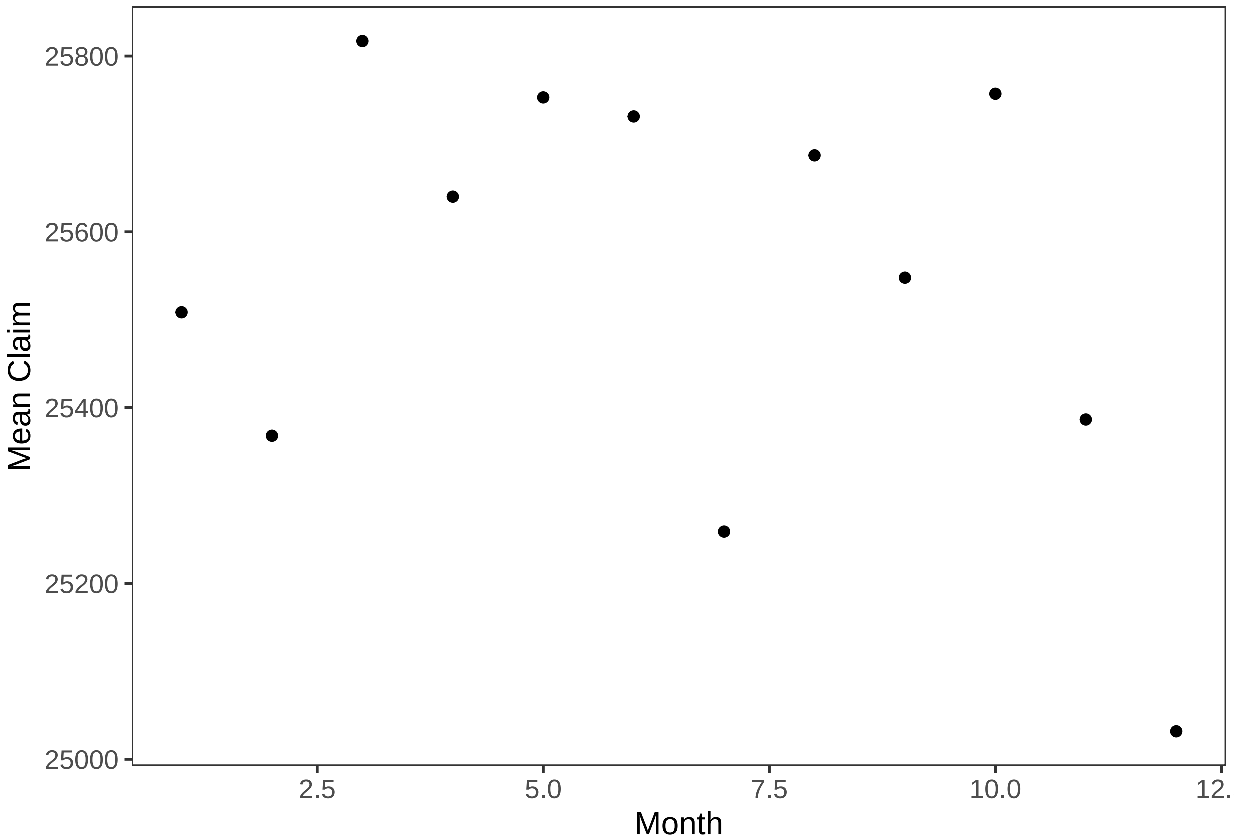

Distribution
Emp
Pois
NegBin
Empirical Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
0
5
10
15
Number of claims
Distribution
Emp
Pois
NegBin
prop
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
Number of claims
Distribution
Emp
Pois
NegBin
Empirical Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Number of claims
Distribution
Emp
Pois
NegBin
prop
0.15
0.10
0.05
0.00
0
5
10
15
20
Number of claims

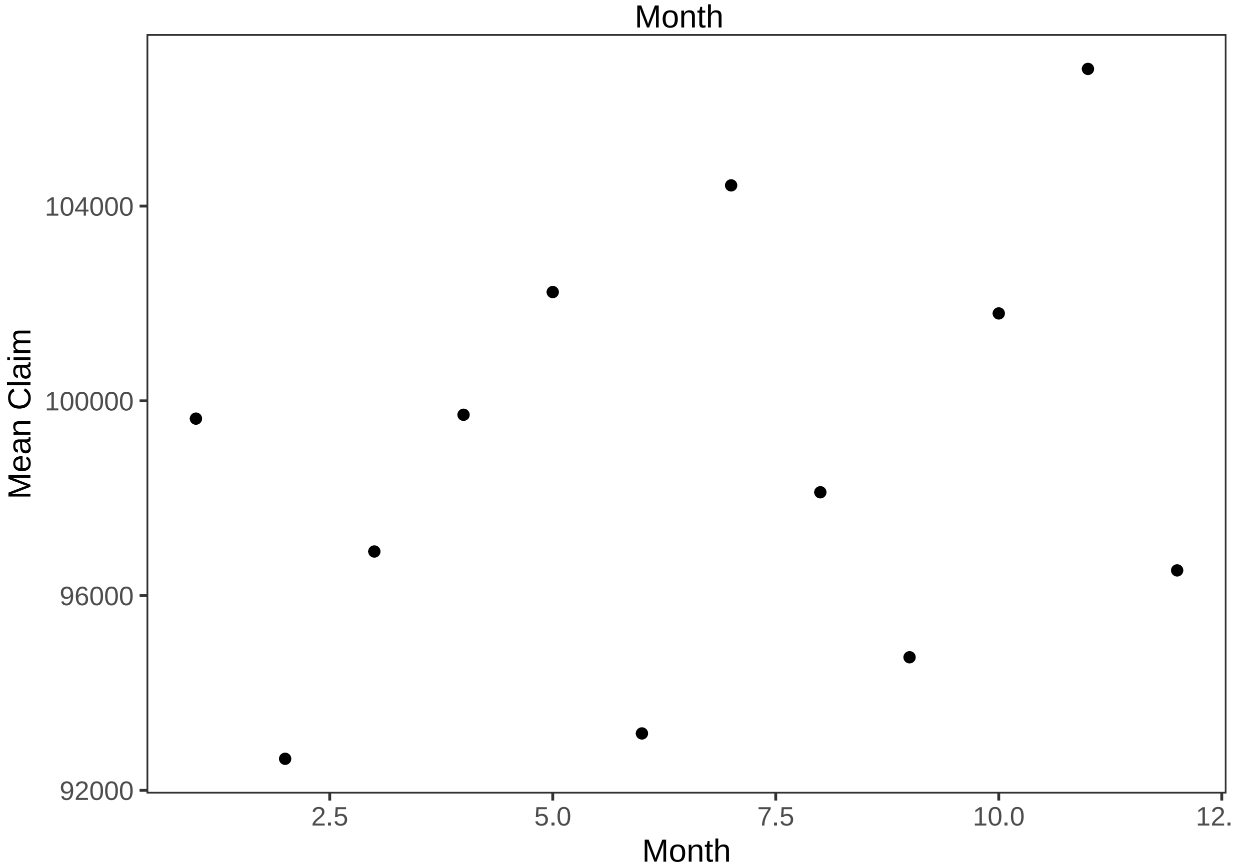




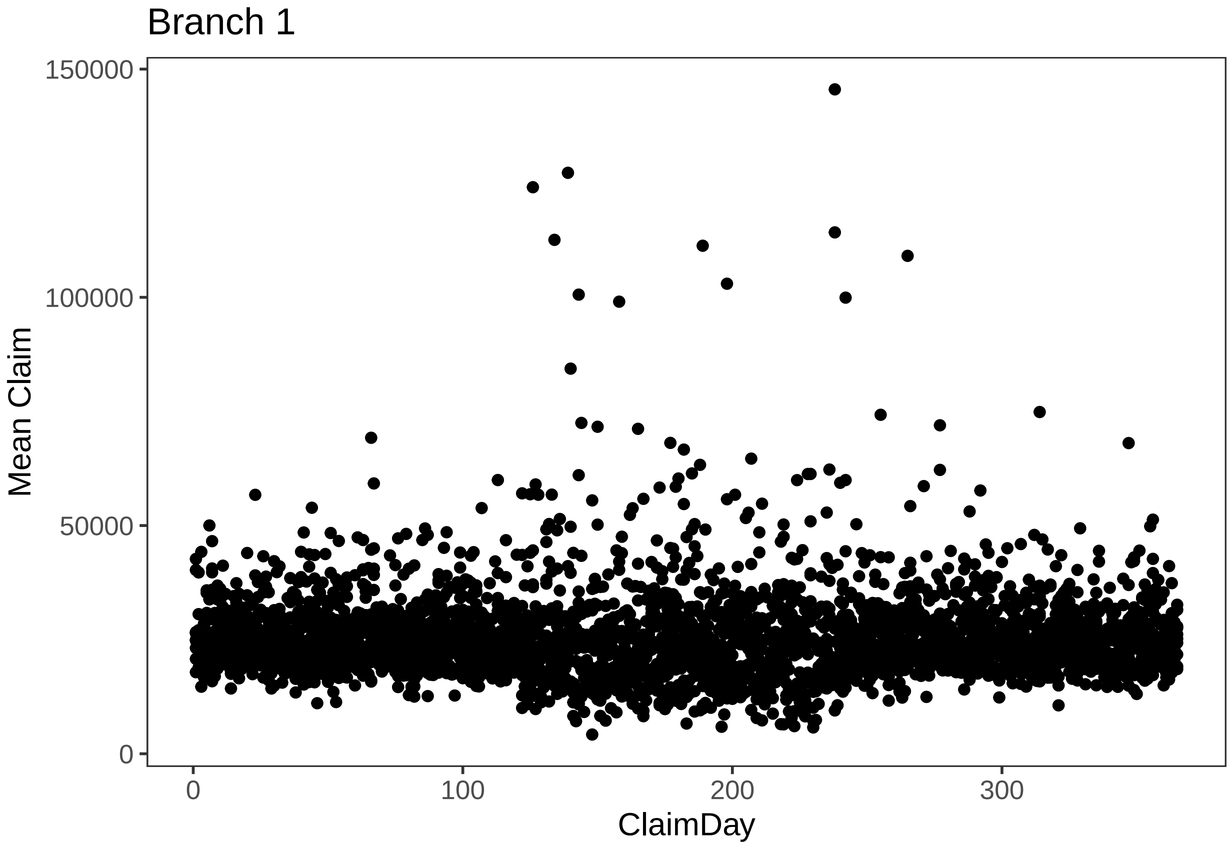
Exercise 2

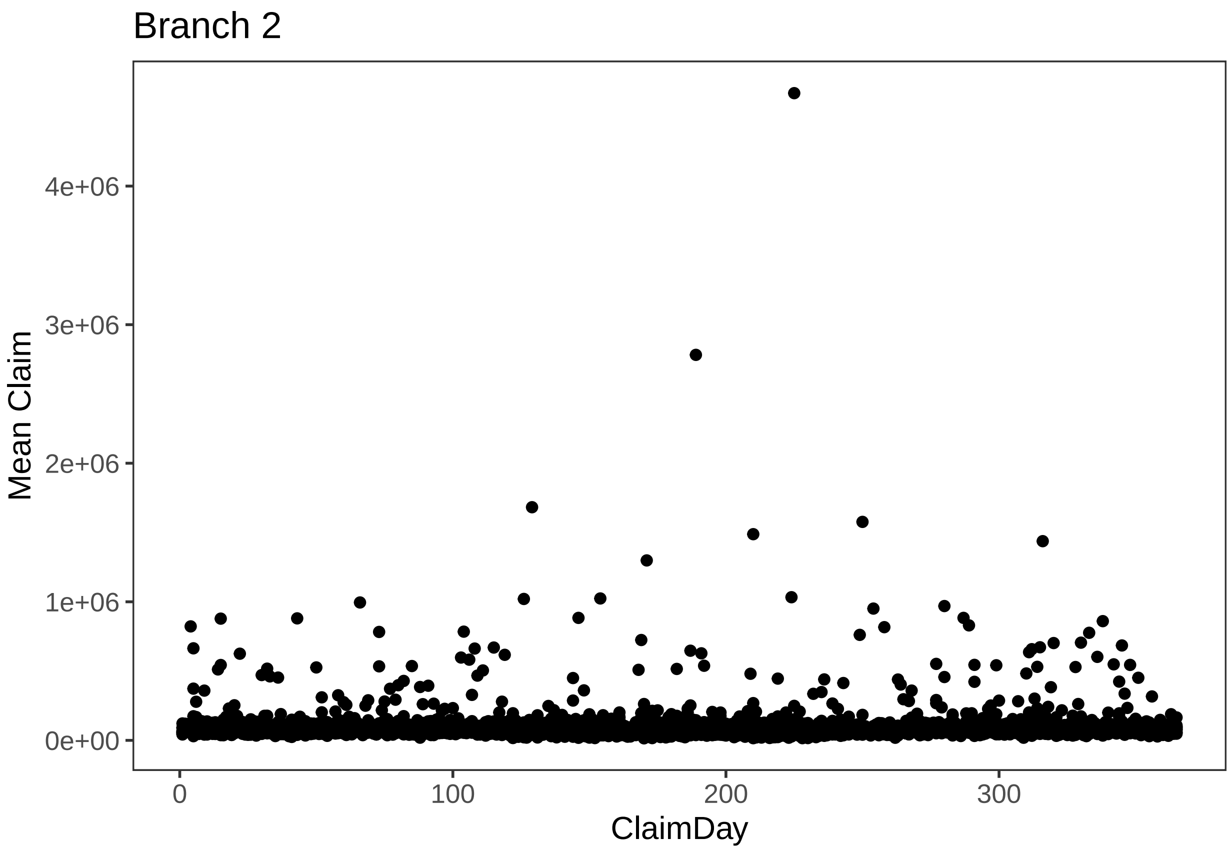
In this exercise we analyze the claim costs over time for each of the two insurance branches. We are looking for possible patterns in the claim costs over time by analyzing the distribution of the claim costs. The following plots look at the average claim cost for each month and insurance branch.



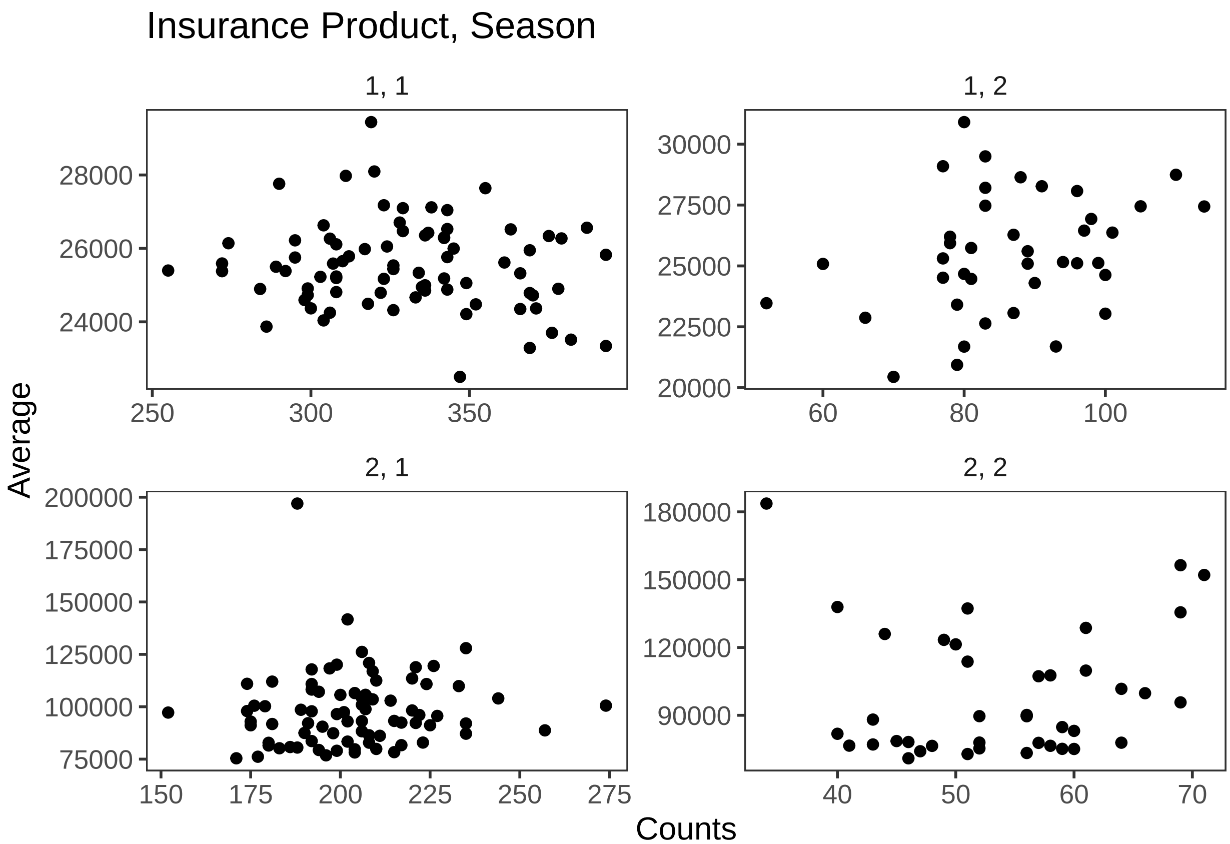


It seems as if the avergae claim costs are time independent, atleast when aggregated on a monthly basis, from the two previous plots. We can also observe the mean of the claim sizes for every day in order to further strengthen the assumption of time independence.



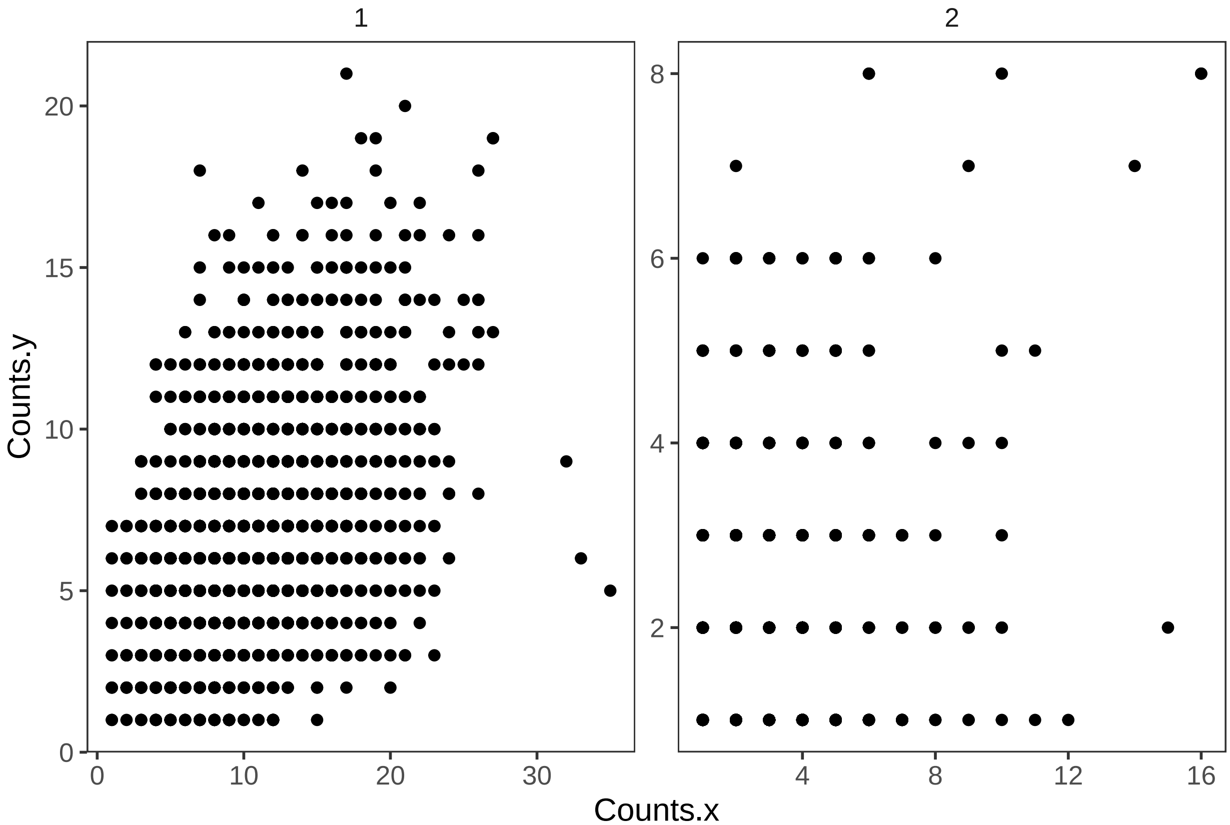


We note from the figures that the average claim sizes seem to be time independet. An interesting factor is however that all extremely deviating claim costs seem to occur during the summer (i.e period 2). In conclussion it seems as if the claim costs are independent of time. Another thing that we want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



From the plots above we cannot see any systematic correlation between claims and claim costs for any of the combinations, meaning that we can model the number of claims and claim costs independantly. We do not observe any time dependencies in terms of average claim costs and there seem to be no systematic correlation between claims and claim costs. In conclussion, the lack of time dependencies for the claim costs means that we can model the claim costs with the assumtion that the claim costs are time independant.

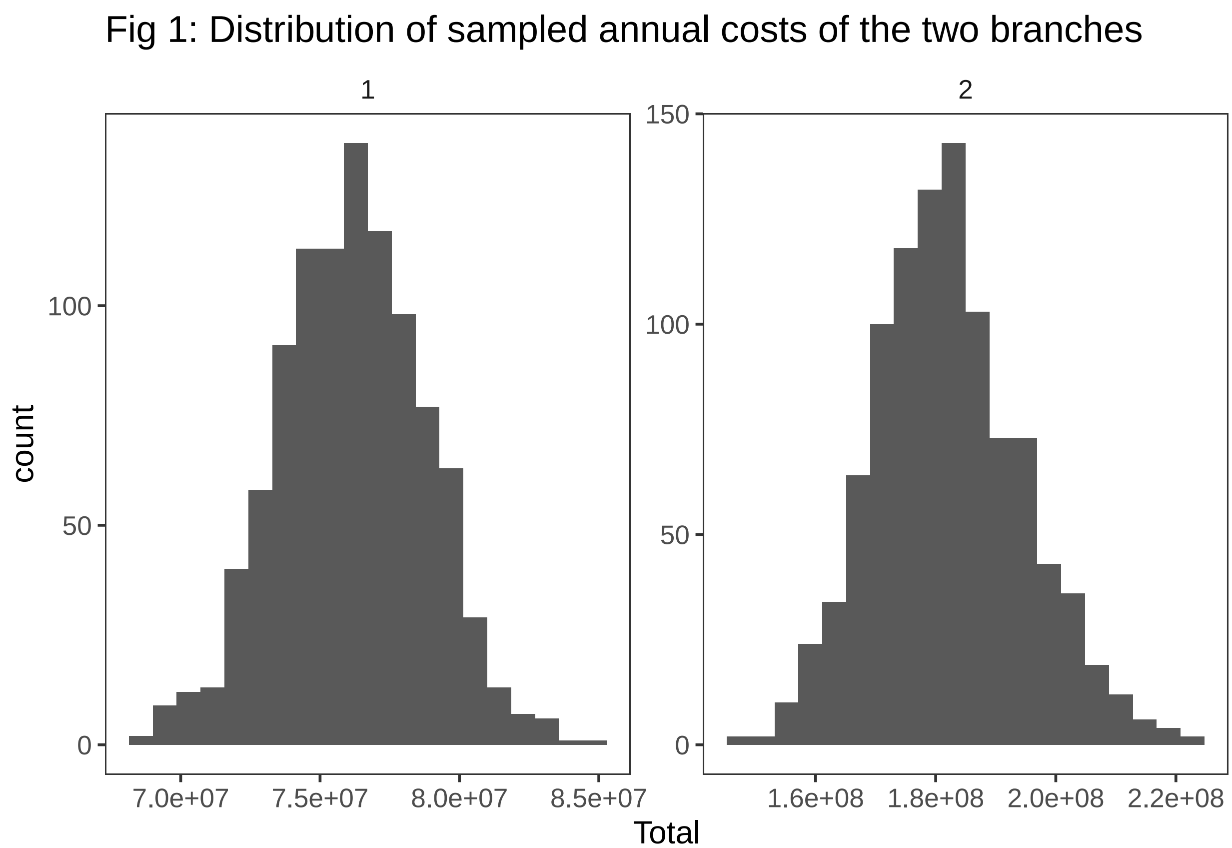
Exercise 3



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. We previosuly mentionded that the claim costs are time independent for both branches as well as that they are independent between the two insurance branches. Due to this we can as previosuly mentioned model claims and the individual claim costs independently.



From Figure 1 above we note that the distributions look fairly normal with branch 2 having both a higher expectation and variance compared to branch 1.

Exercise 5

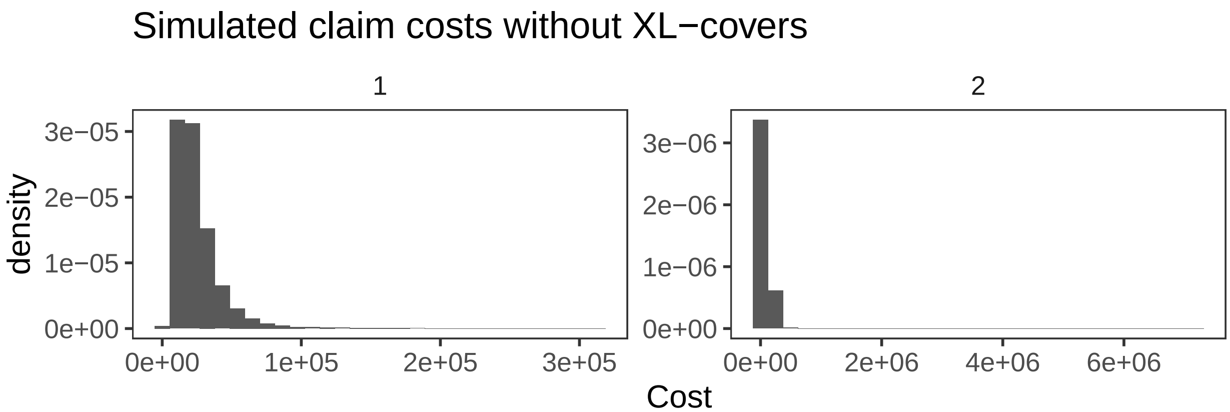
We now want to implement two separate XL-covers. The covers caps losses for the 10% worst claims for each respective insurance branch. For our two branches these cut-off M_1 and M_2 are the following

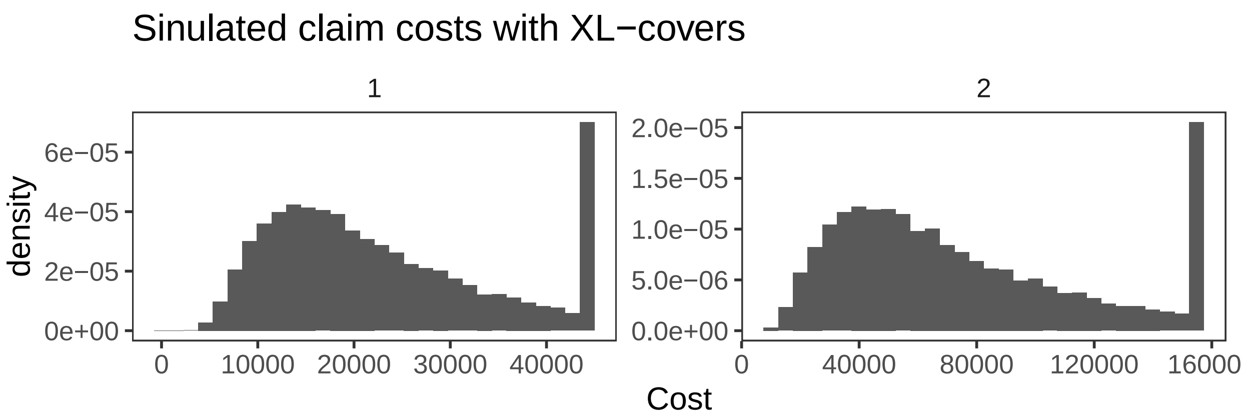
Table 1: Cut-offs

Type	M
1	44562
2	154011

We note that these cutoffs are vastly different for the two branches which of course is to be expected since they likely insure against different things. Graphically the distribution of the individual claim costs looks as follows when, and when not, taking the XL-covers into account.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



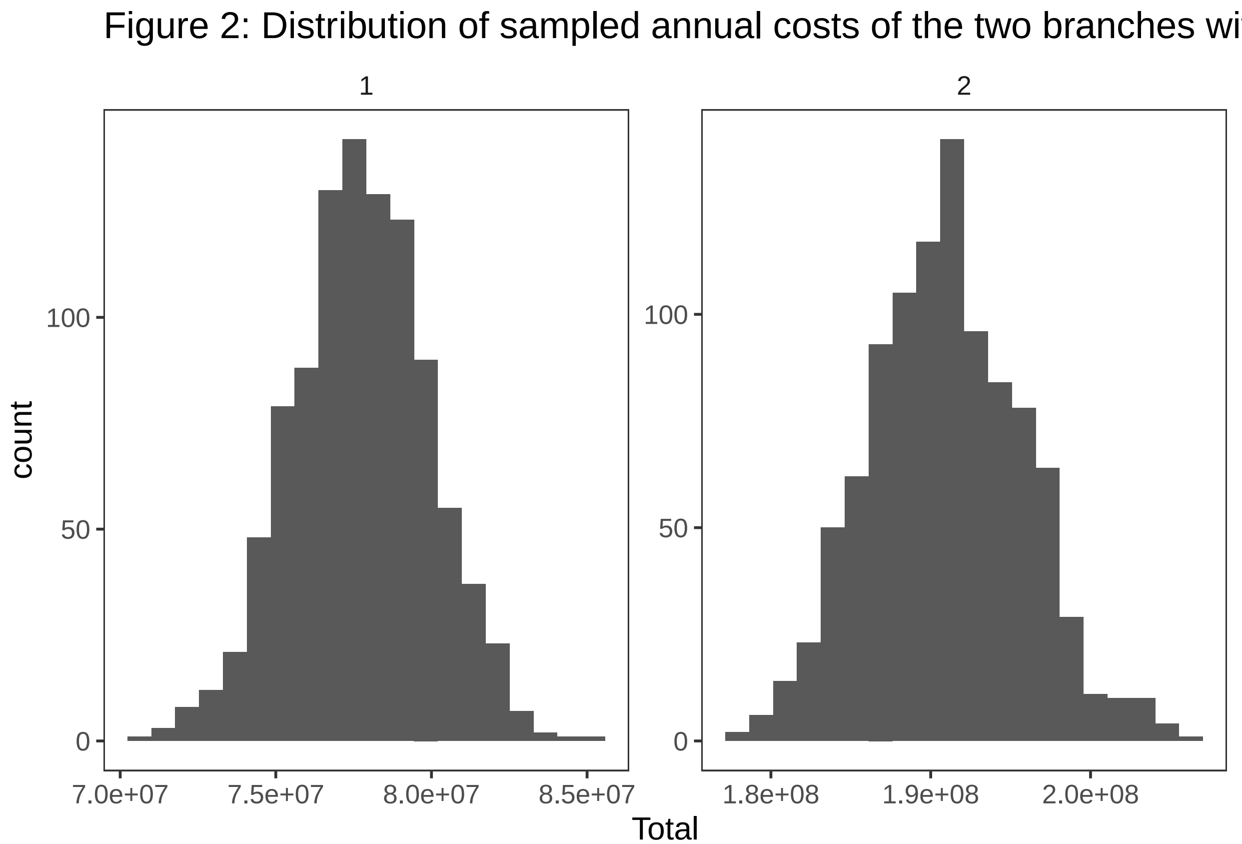


From the figures above we note that both XL-covers seem to cover some very deviating claim costs, specifically in the case of insurance branch 2 which in some sense could be expected due to the higher variance compared to branch 1. Due to the possibility of these large outliers the cost of the XL-covers will be very high.

Table 2: Prices for the two XL-covers

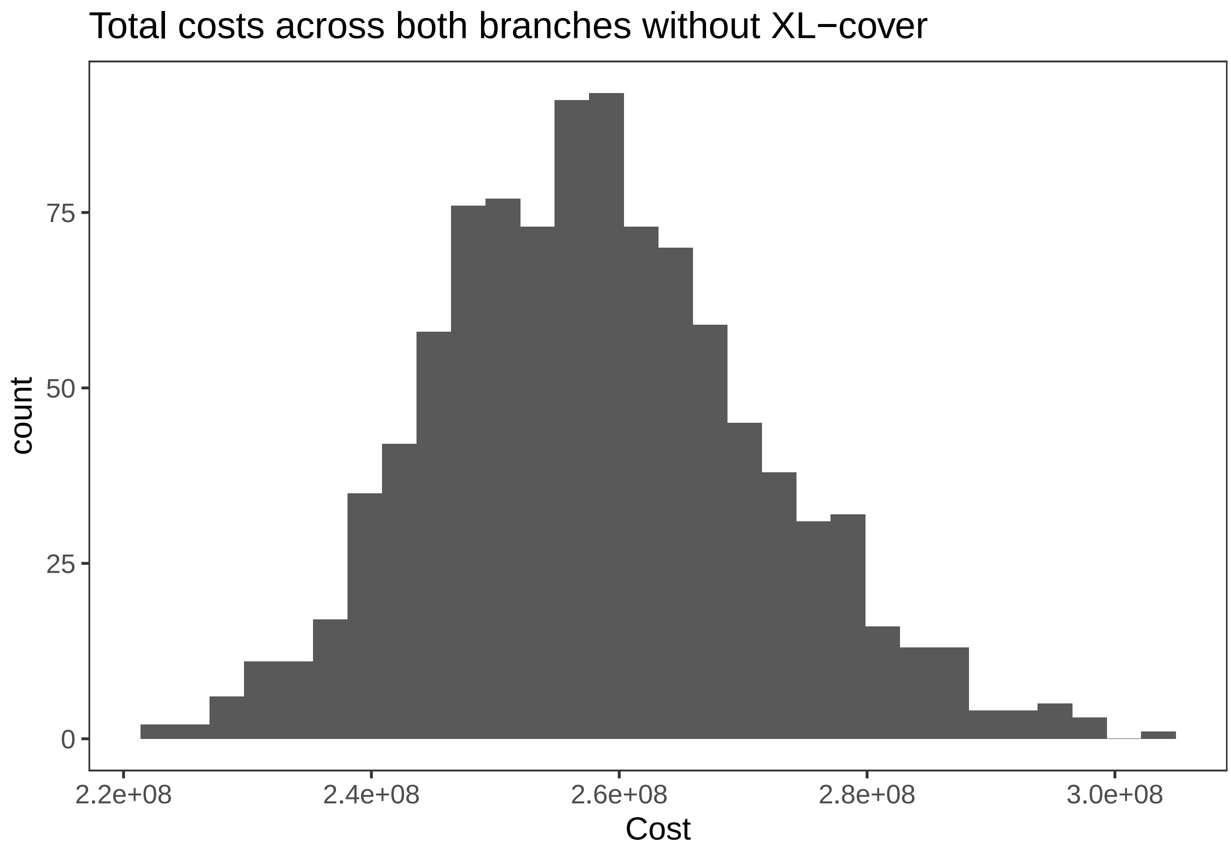
Type	Price
1	9473202
2	53191862

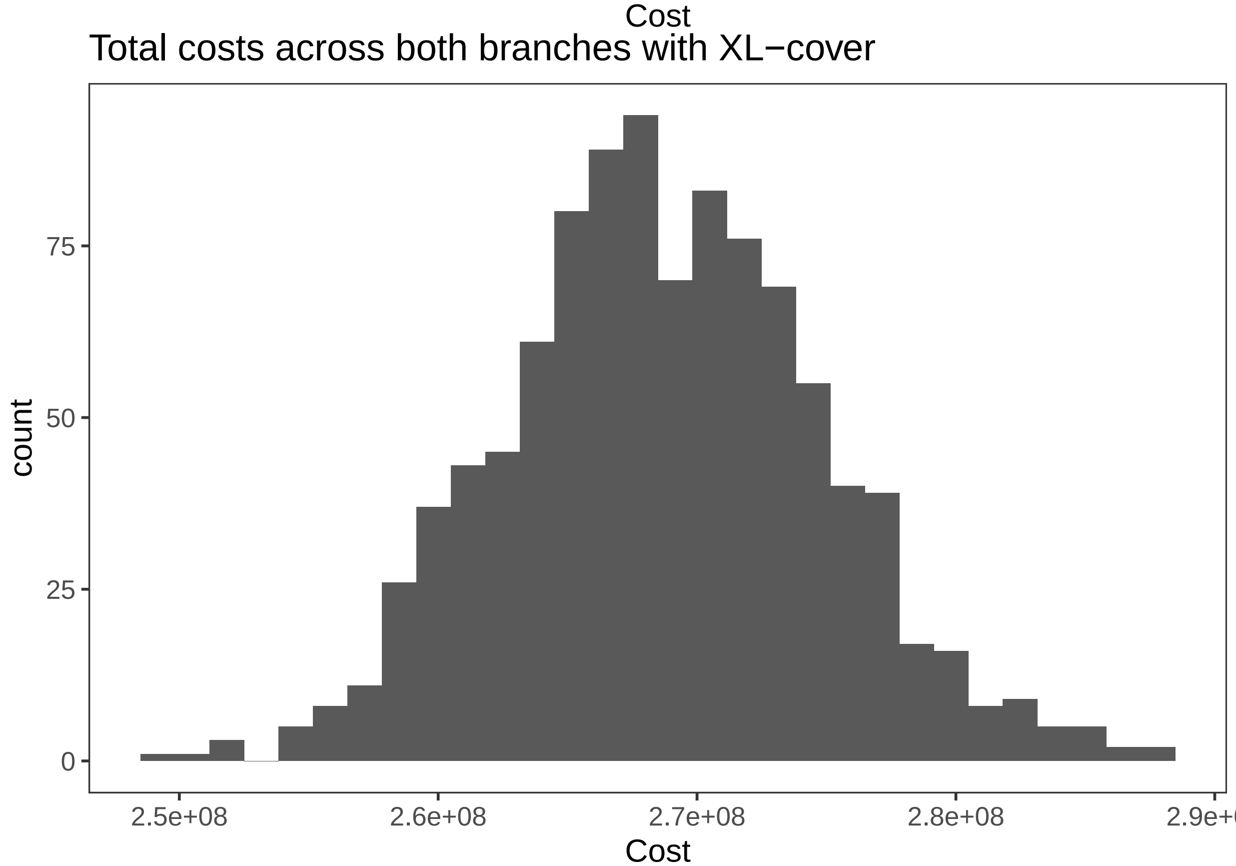
As previosuly mentioned both covers turn out to be fairly pricy, specifically the cover for branch 2. We now want to examine how the purchase of these XL-covers would impact our total losses for the forthcoming year. We saw a simulated distribution of the losses for the two branches without the XL-covers in Figure 1. If we implement the covers we get the following simulated distributions.



When comparing Figure 1 to Figure 2 we note that that the approximate expected cost for the forthcoming year is slighly higher when purchasing the XL-cover for the respective branches. This increase in the expected cost does however come with the benefit of reducing the weight of the tails of the distributions. As such, the purchase of the XL-covers seem to increase the average cost whilst reducing the overall risk. For a risk averse insurance provide these XL-covers could therefore serve as an excellent tool to reduce the overall risk.

We can also observe the how the purchase of both covers affect the distribution of the overall annual cost.



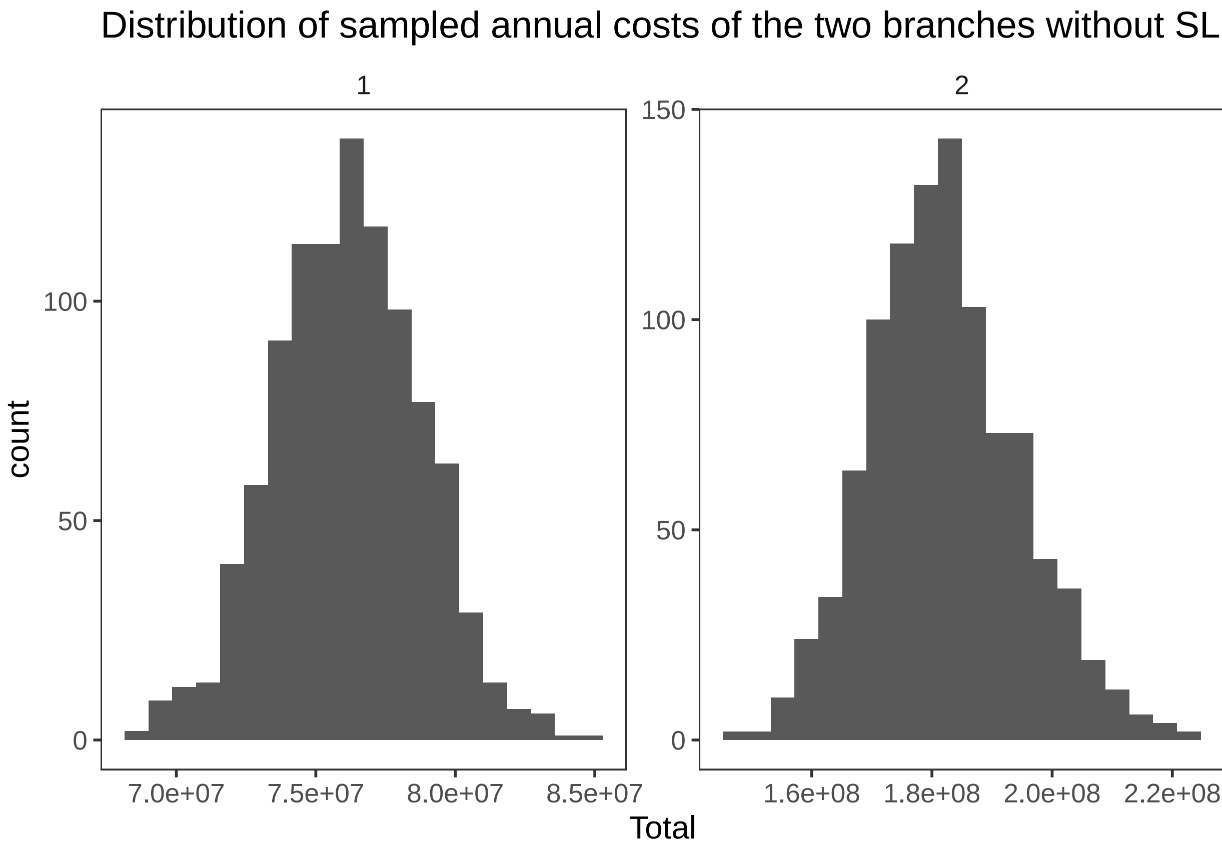


Once again we see that the distribution is lighter-tailed whilst having a higher expectation.

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. We can once again plot the simulated annual costs with and without the SL-covers.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



In the case of the SL-covers we note that, unlike for the XL-covers, that we have no extreme deviations in the annual cost (compared to the deviations in the distribution of the claim costs). This is due to the fact that the annual cost is the aggregation of a very large number of individual claim costs, thus an outlier of the same proportions as those in the case of the individual claim costs will be theoretically extremely unlikely and practically never occur. In other words, by aggregating the claims we reduce the overall risk of large deviations. As such the prices for the SL-covers should in some sense be proportionally lower compared to those of the XL-covers.

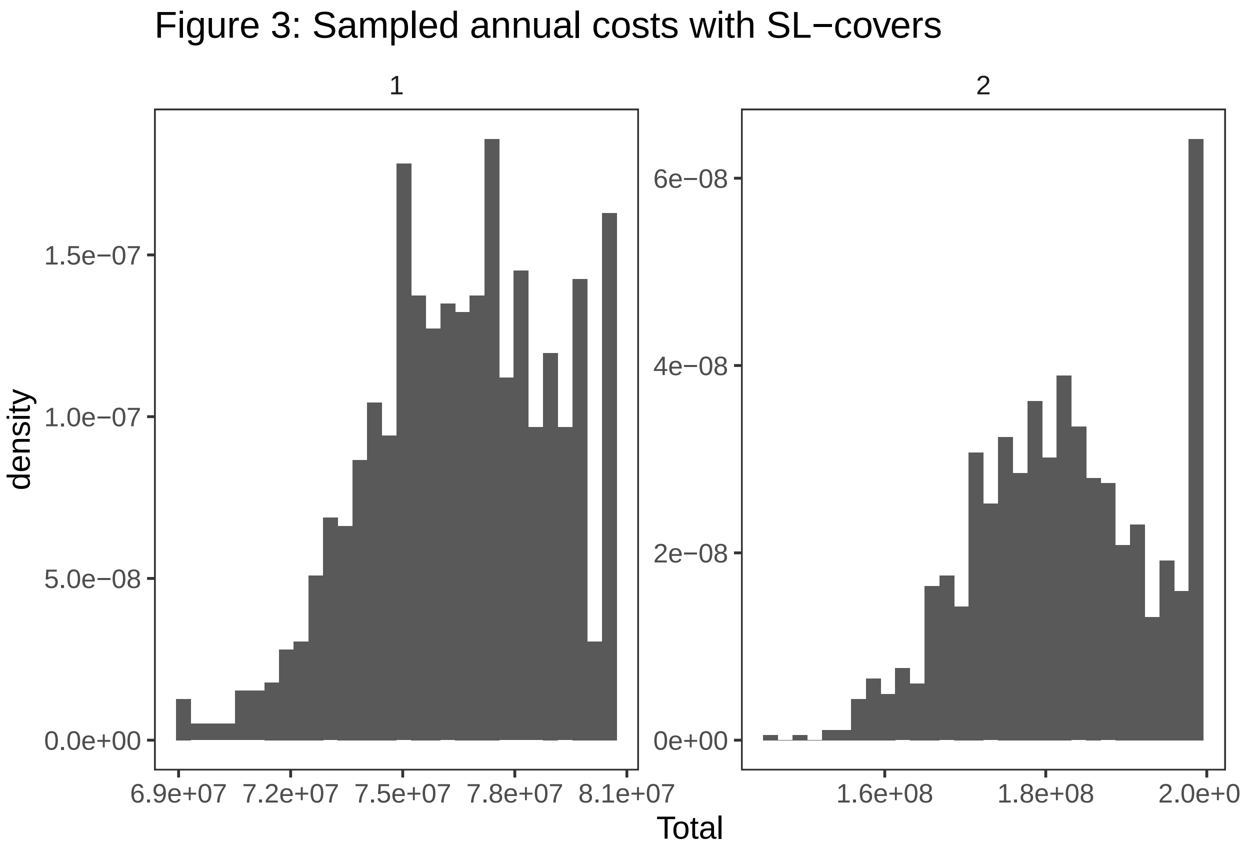
The prices of the two SL-covers are as follows

Table 3: Prices for the two SL-covers

Type	Price
1	136864.5
2	839111.3

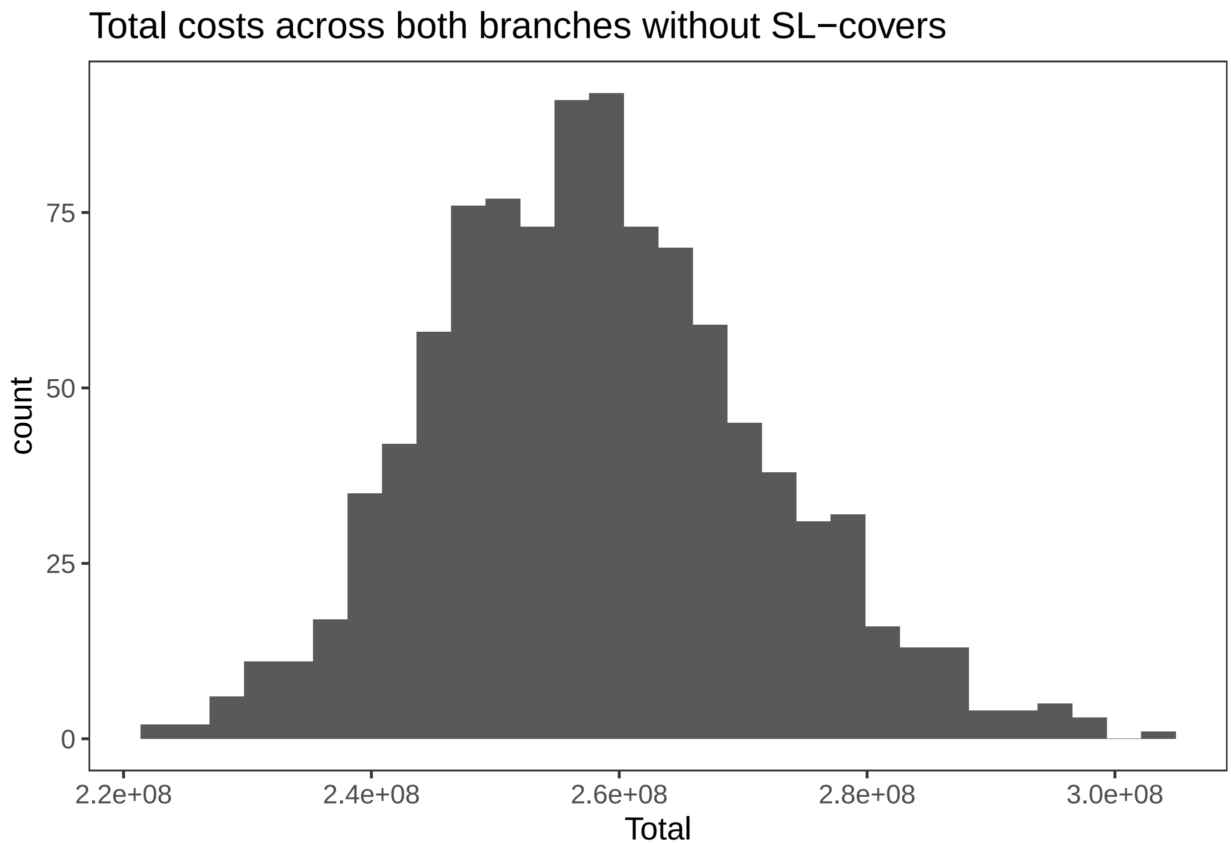
Here we note that the prices are actually extremely low, for insurance branch 1 the annual cost of the SL-cover but a drop in the ocean compared to the overall annual claim costs. The price for the SL-cover for branch 2 is a bit more pricy which is due to the fact of the higher variance of the annual claim costs distribution as was previously noted.

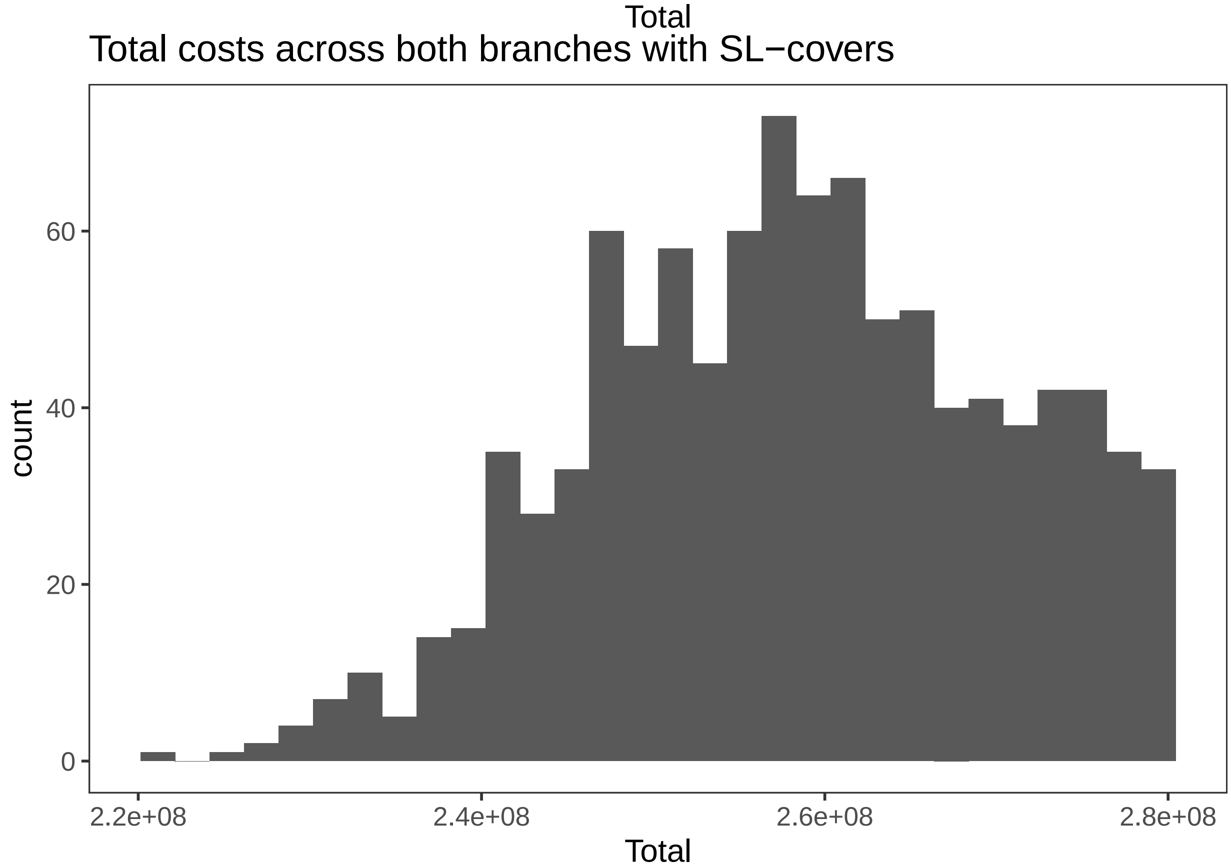
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



When comparing Figure 1 and Figure 3 we note that the expected cost is slightly higher due to the price of the SL-cover but more importantly that we have cut-offs which essentially prohibits any larger deviations in the overall cost past the cut-off point. By doing this the distributions for the overall costs are skewed slightly positiveley.

We can also examine how the purchase of both covers impact the overall annual claim costs.



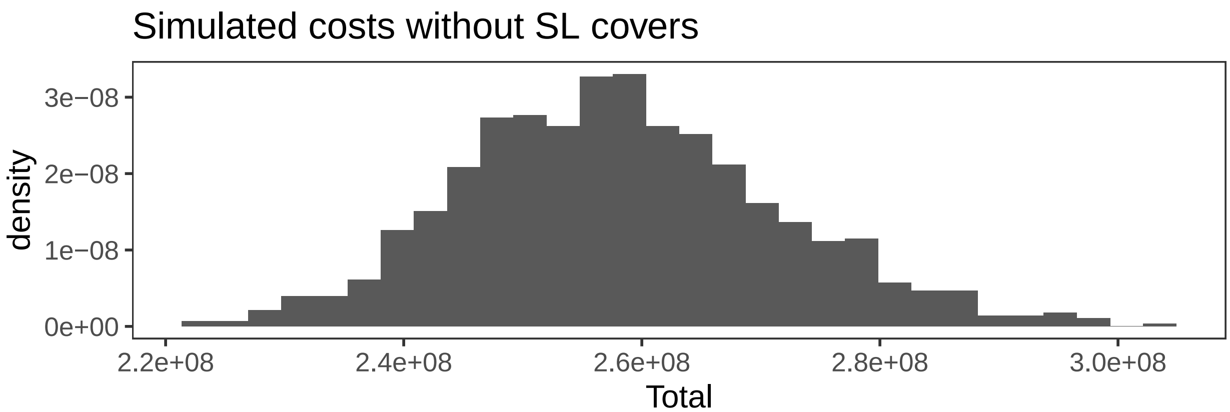


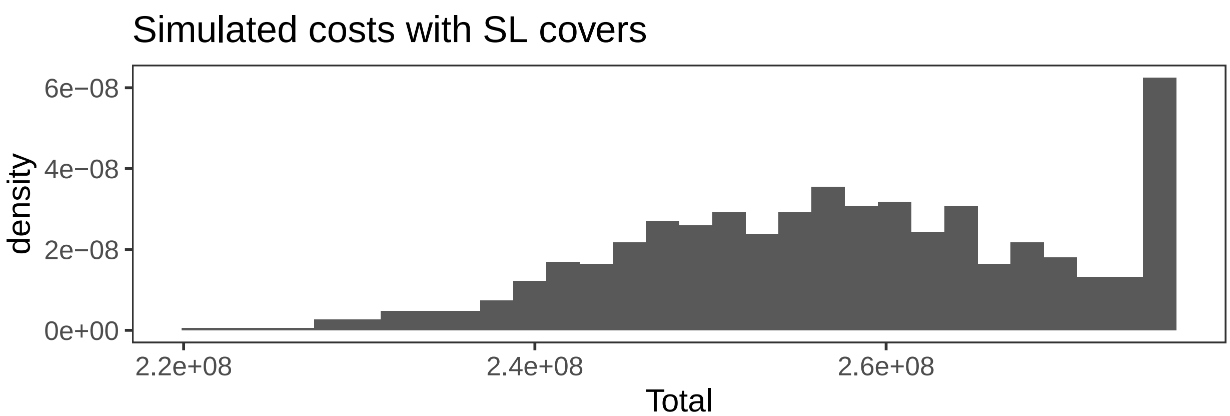
By purchasing both covers we increase the expectation slightly whilst also skewing the distribution positivley as in the case of the individual cost distributions of the branches.

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.

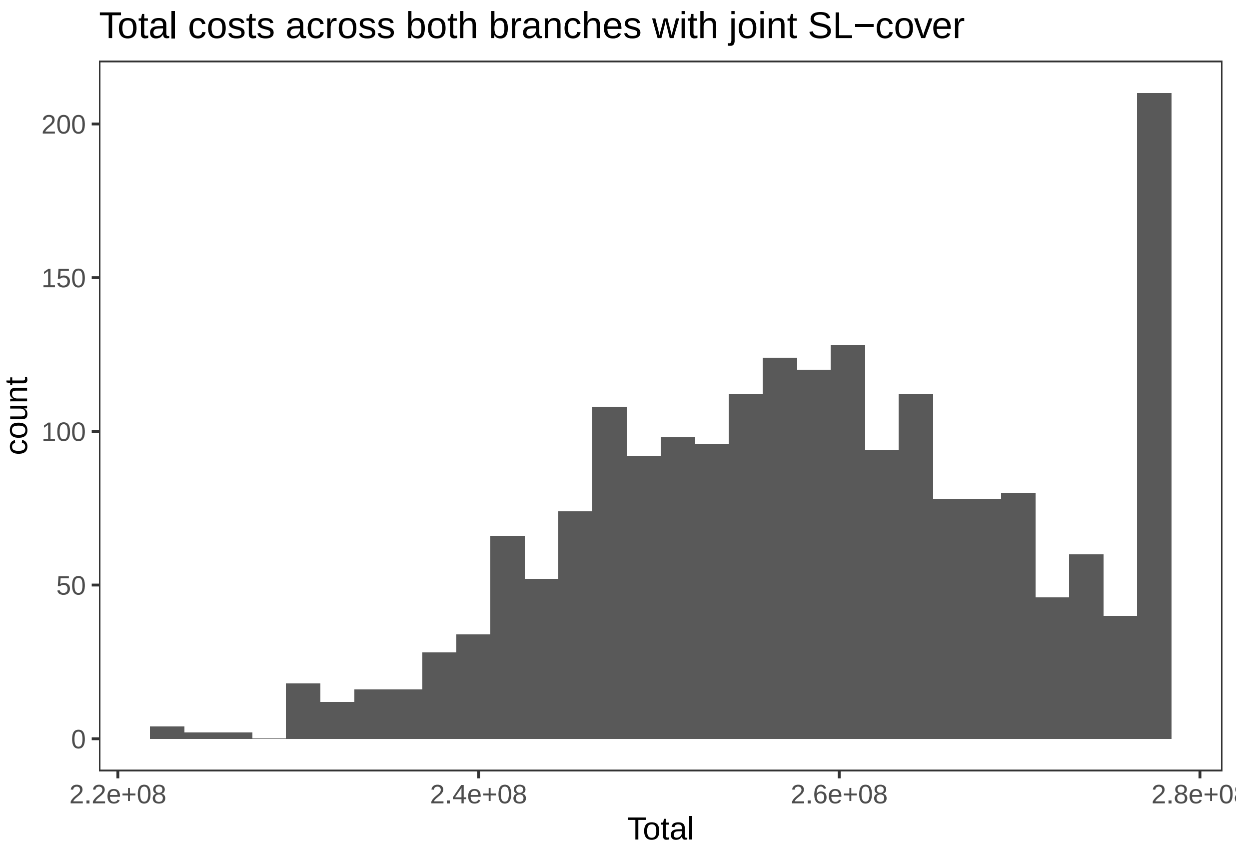
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```





Like for the previous SL-covers we note that the distribution of the total annual cost without the purchase of the cover lacks any extreme outliers and as a result the overall cost for the SL-insurer and as a consequence the price will be lower.

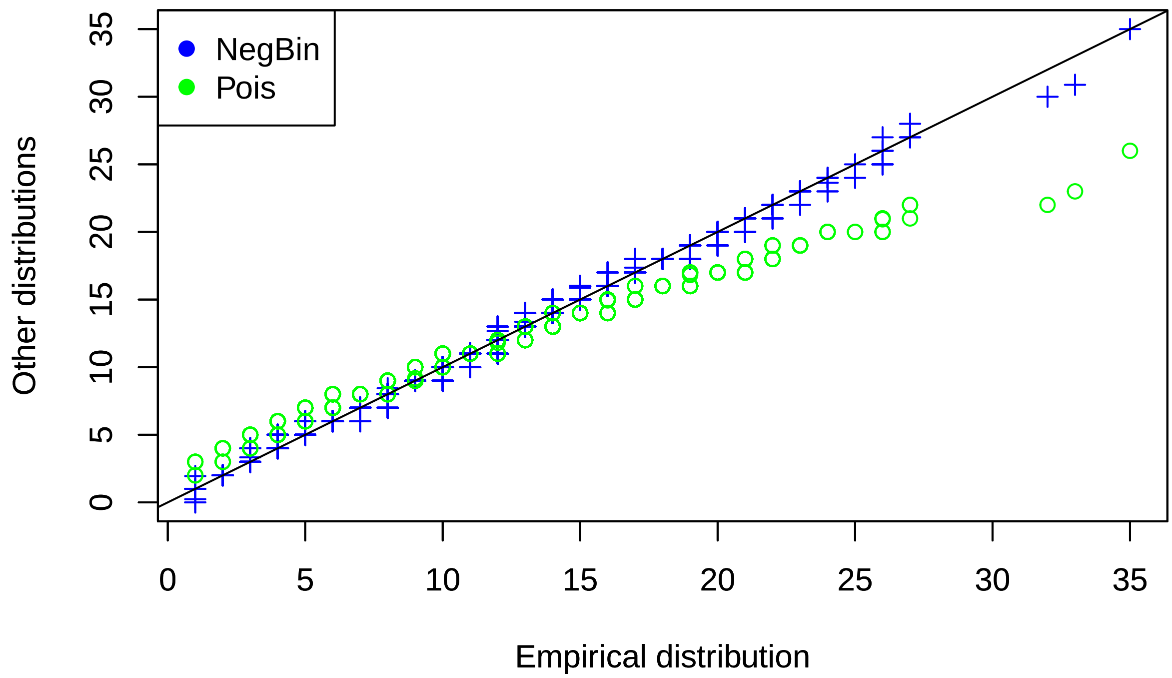
The cost of this cover, based on simulated data, is 8.3657639×10^5 which is once again lower than than the previously examined covers, this is because we have aggregated on yet another level which has further reduced the risk.



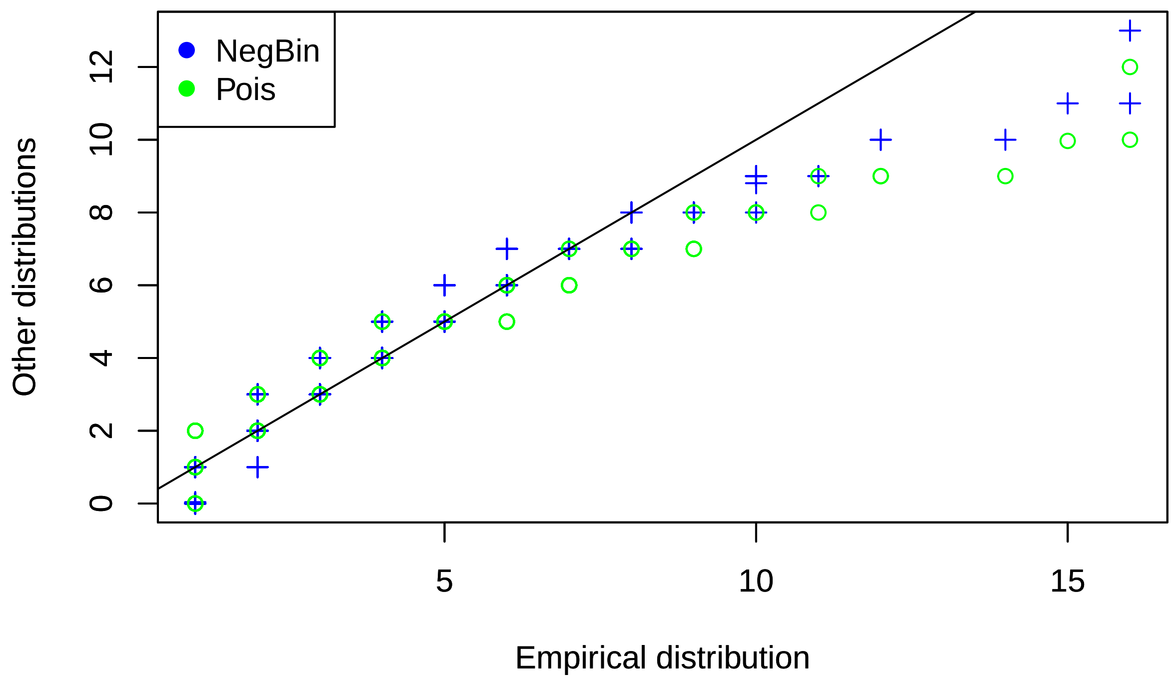
We see from the two histograms that the total SL-cover implies a higher expected cost but also limits the total annual cost to $M = 2.7615045 \times 10^8$. As such a more risk averse insurance provider might opt for the SL-cover to completly exclude the possiblity of large deviations in the annual costs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

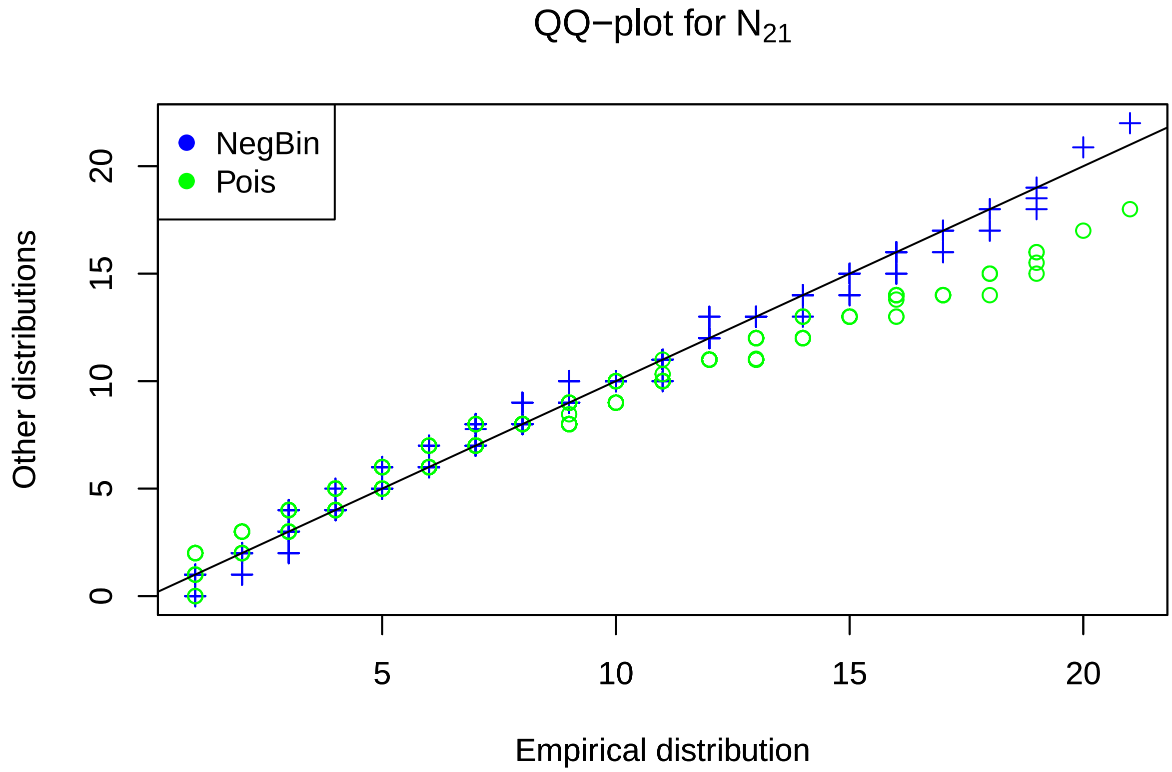
Appendix

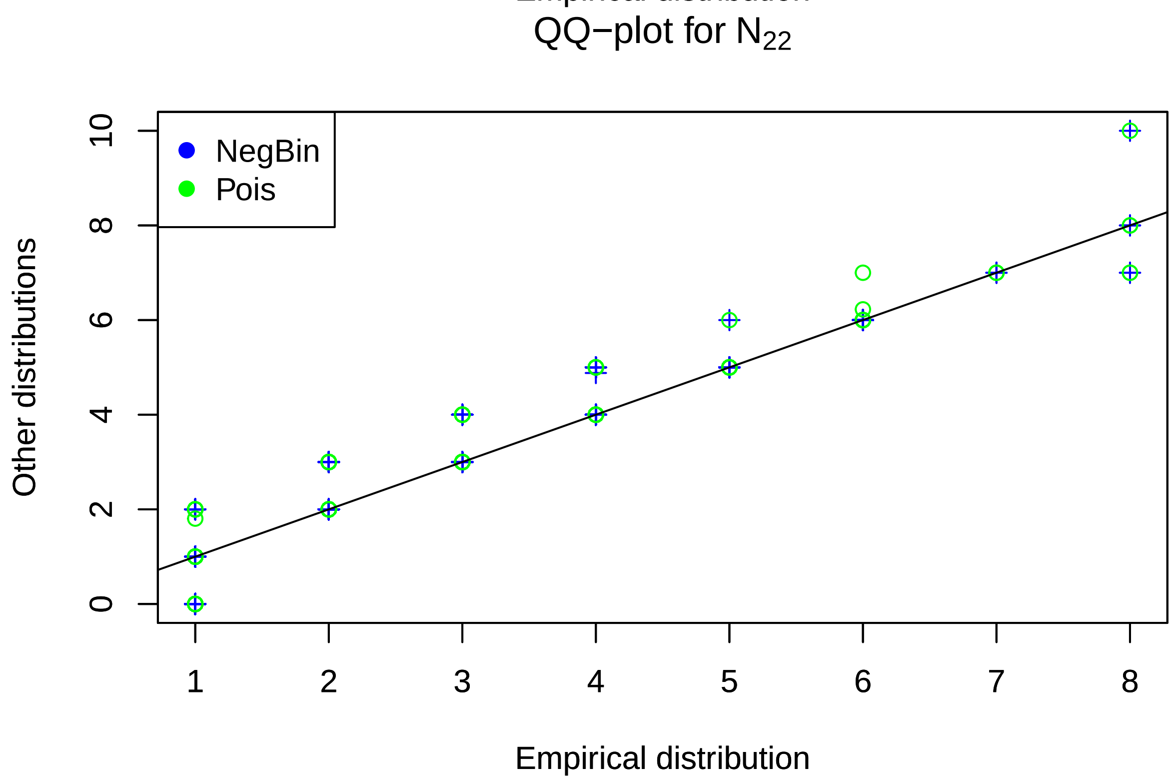
QQ−plot for N_{11}



QQ−plot for N_{12}




QQ−plot for N21
NegBin
Pois
Other distributions
Empirical distribution
0
5
10
15
20


QQ−plot for N22
NegBin
Pois
Other distributions
Empirical distribution
0
2
4
6
8
10
1
2
3
4
5
6
7
8