MT7027: Project 1

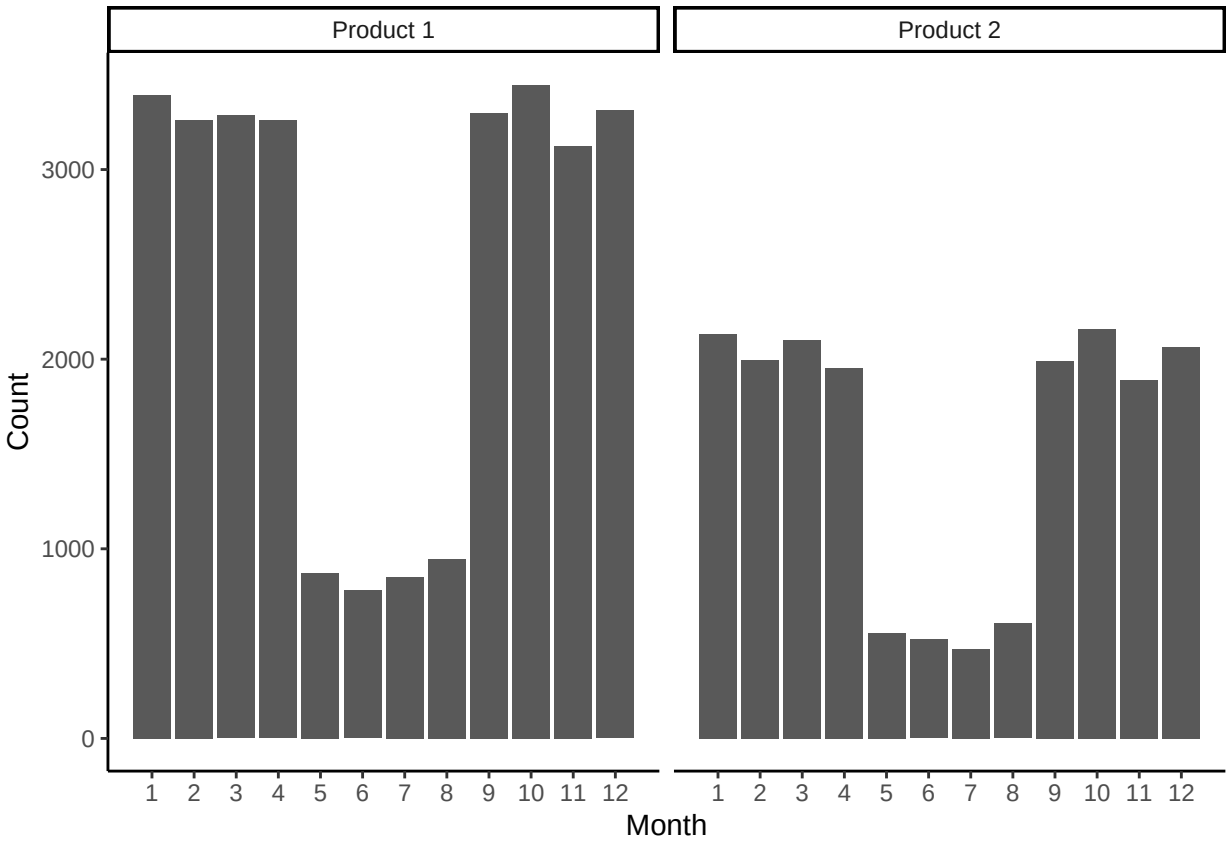
Anton Stråhle, Jan Alexandersson & Max Sjödin

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.

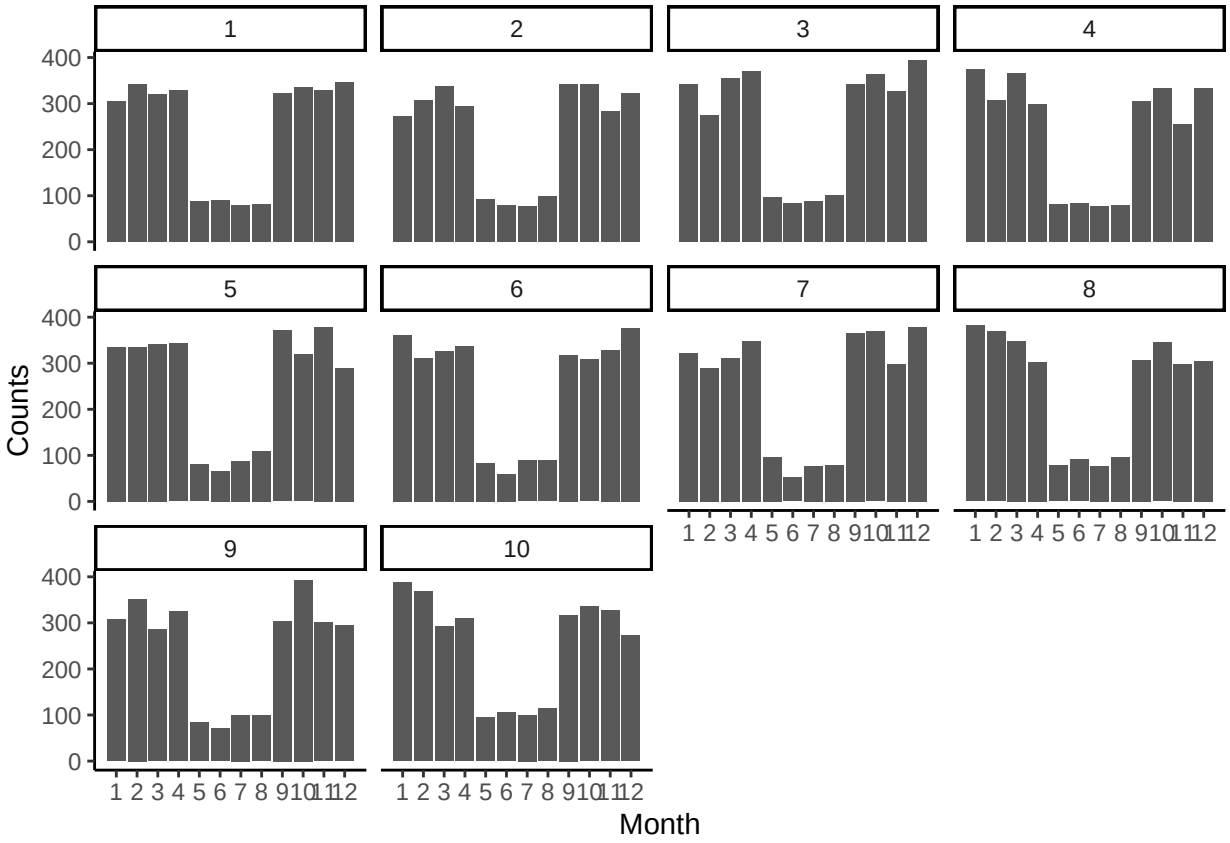
Exercise 1

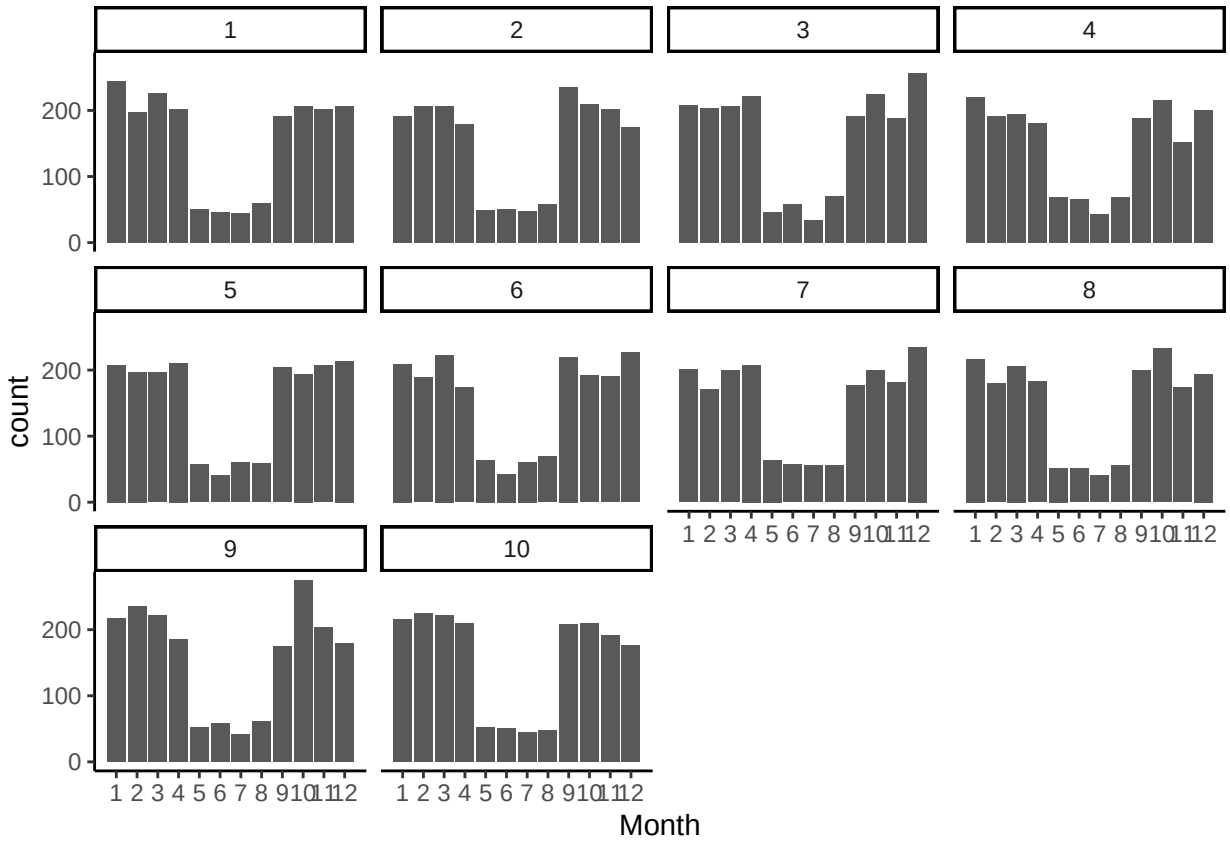
In this exercsie we want to find trends in the data for the two insurance branches in order to model future claims in a block-wise manner. Since the data is structured in a way such that we only have the claim day (i.e. $1, 2, ..., 3650$) we assume that 365 day/year and that a month is 365/12 days (to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



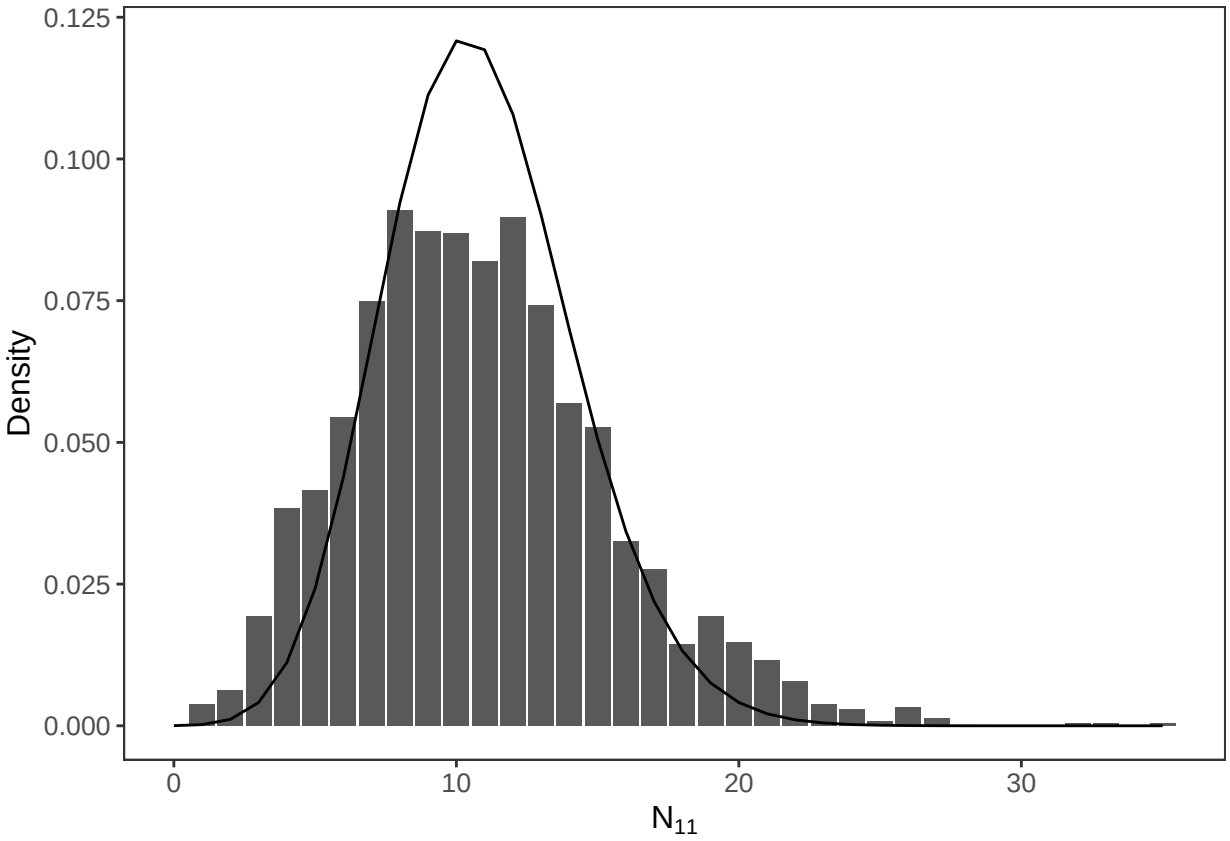


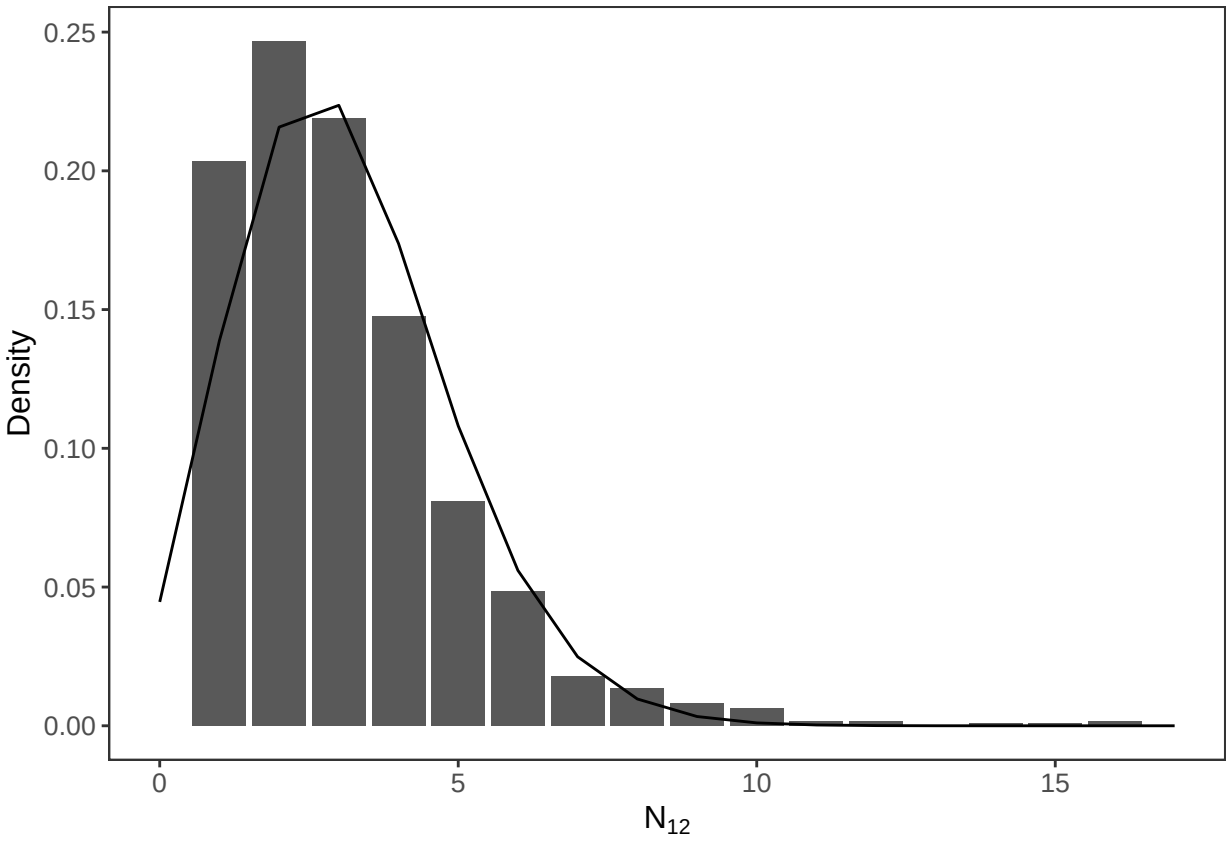

```
## $y
## [1] "Counts"
##
## attr(,"class")
## [1] "labels"
```

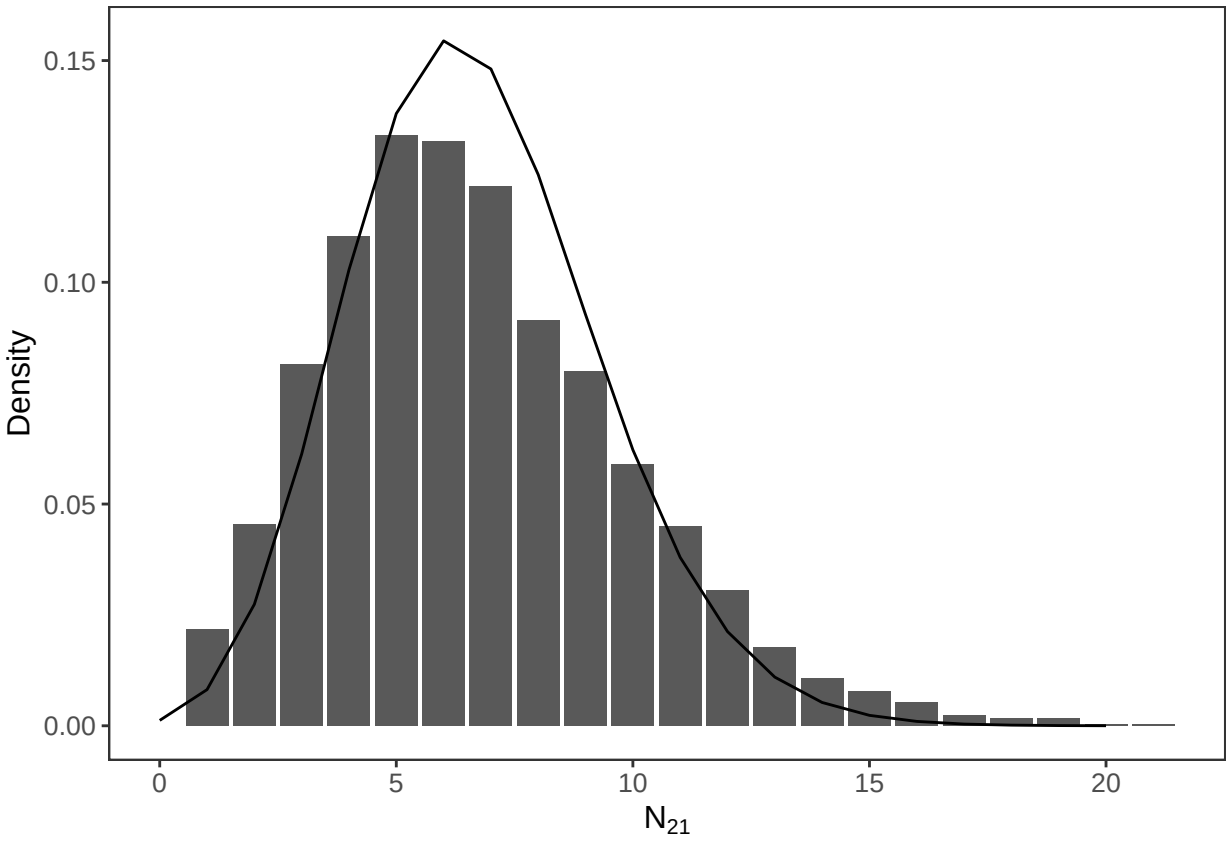
We note from the previous plots that there does not seem to be any major difference in the number of claims between the years.

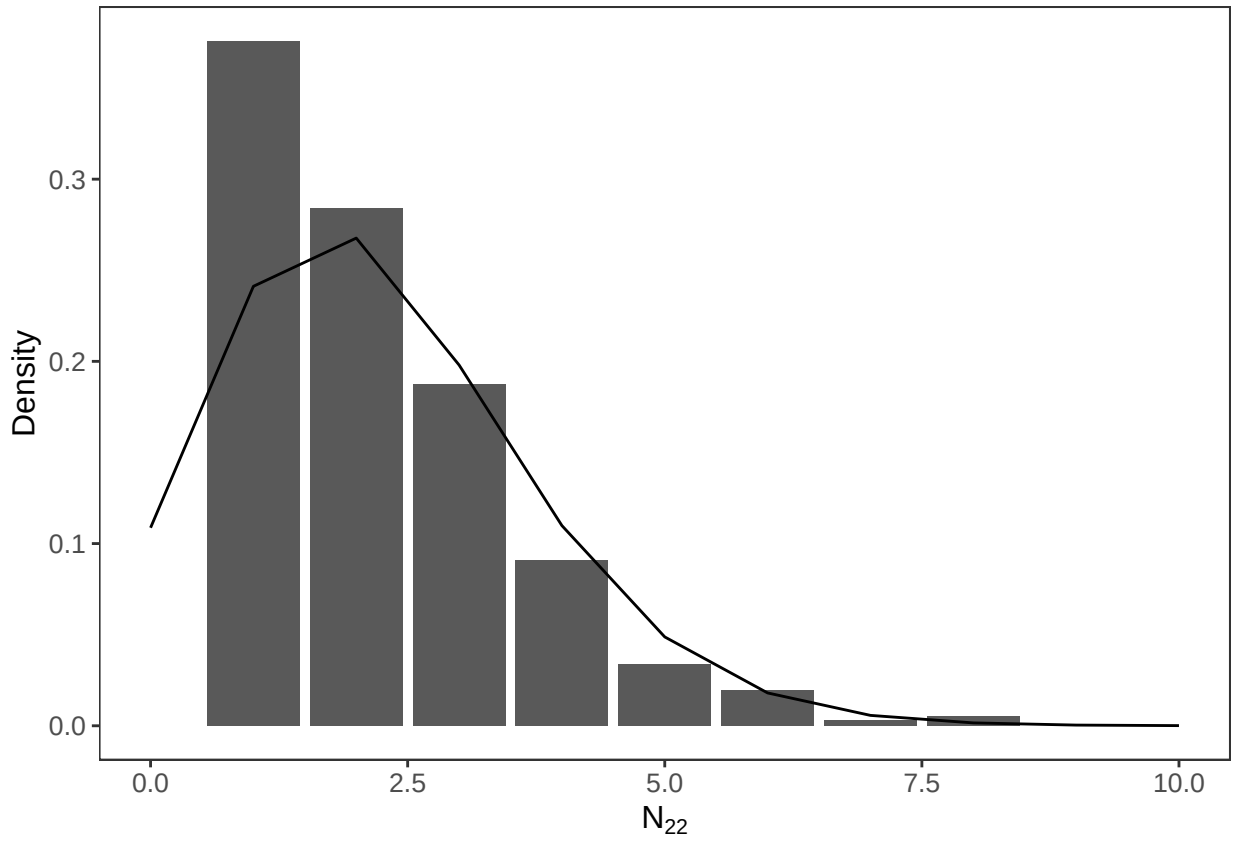
We now wish to fit homogeneous Poisson distributions N_ij (where i represents the insurance type and j represents the season) to the months of each season and each insurance product.

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


0.125
0.100
0.075
0.050
0.025
0.000
Density
0
10
20
30
N11


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

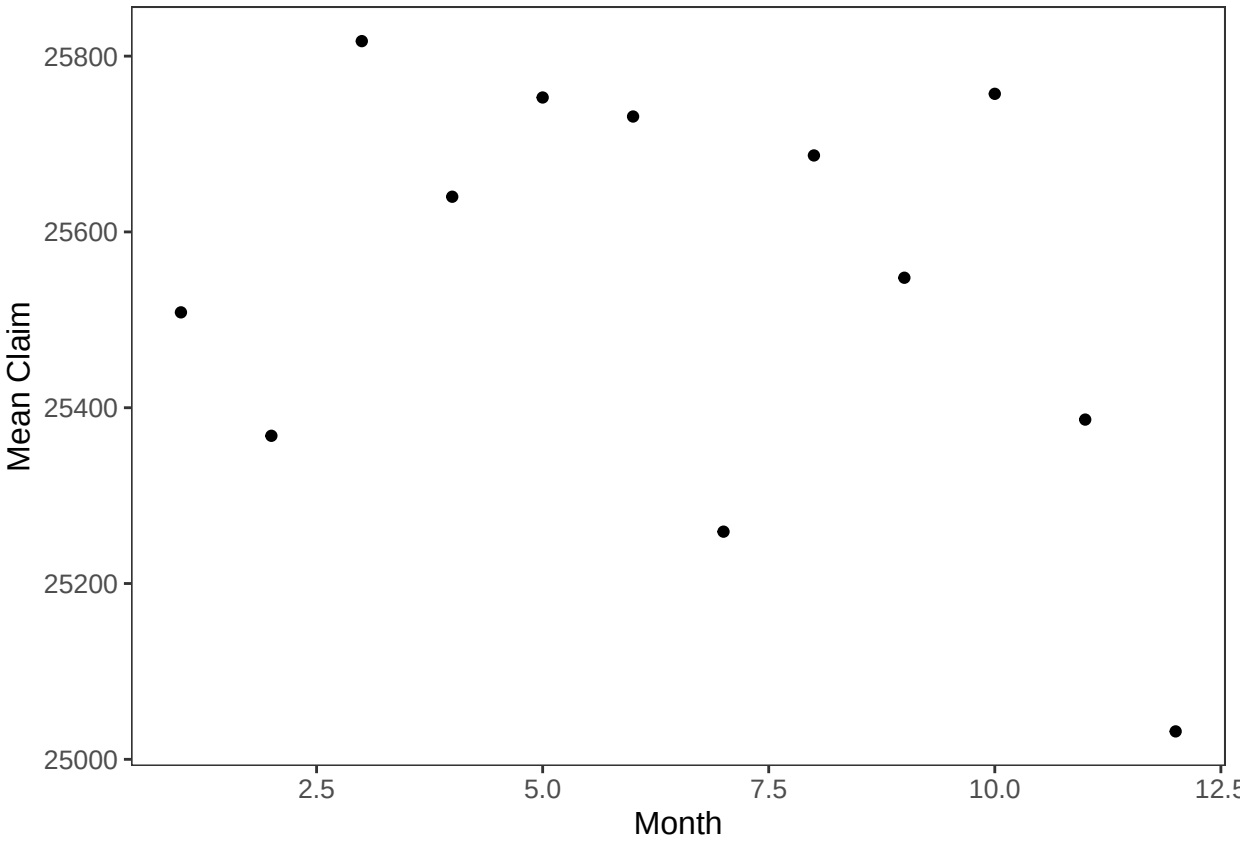

0.15
0.10
0.05
0.00
Density
0
5
10
15
20
N21

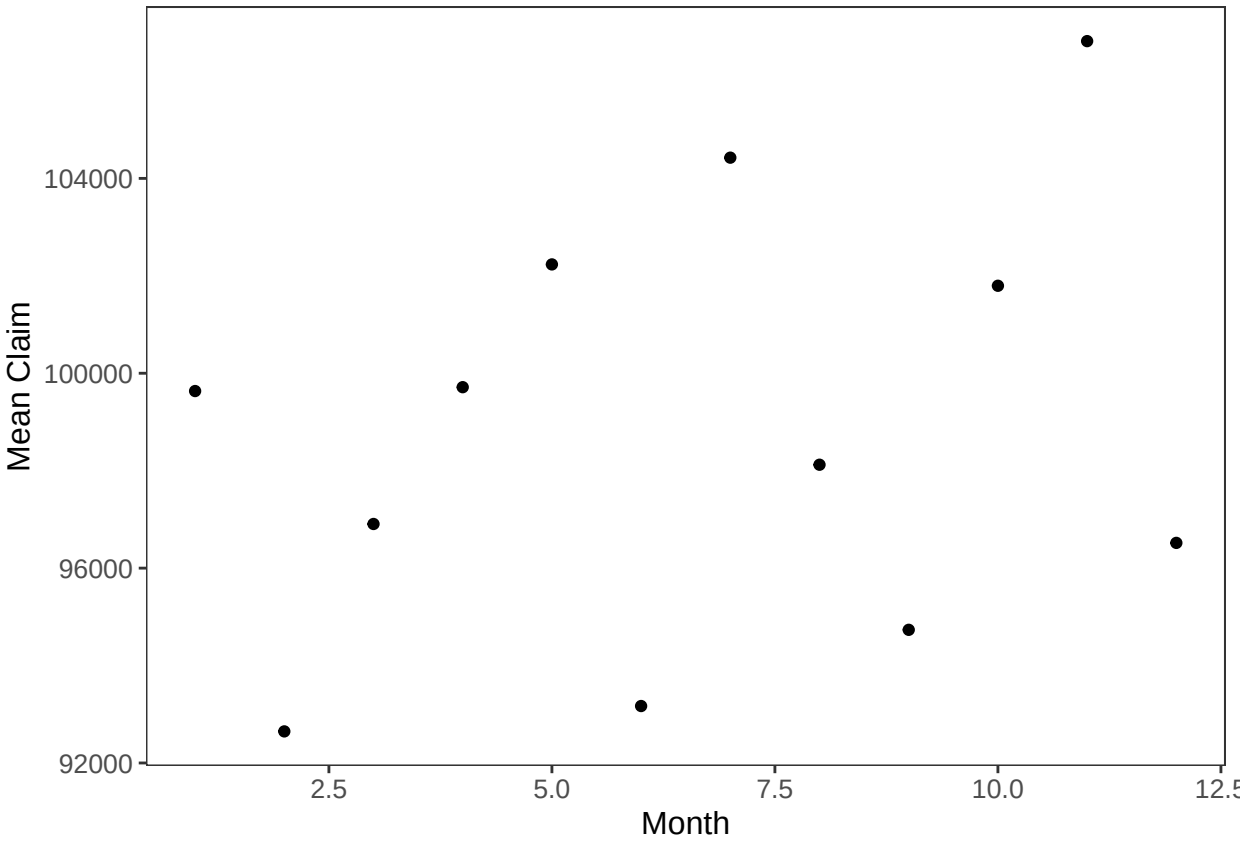


The Poisson variables which have been fit for the different seasons and insurance branches have had their parameter λ_{ij} estimated through maximum-likelihood methods using the data from the 10 previous years. We note that we seem to have overdispersion for N_{11} and N_{12}, meaning that the variance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

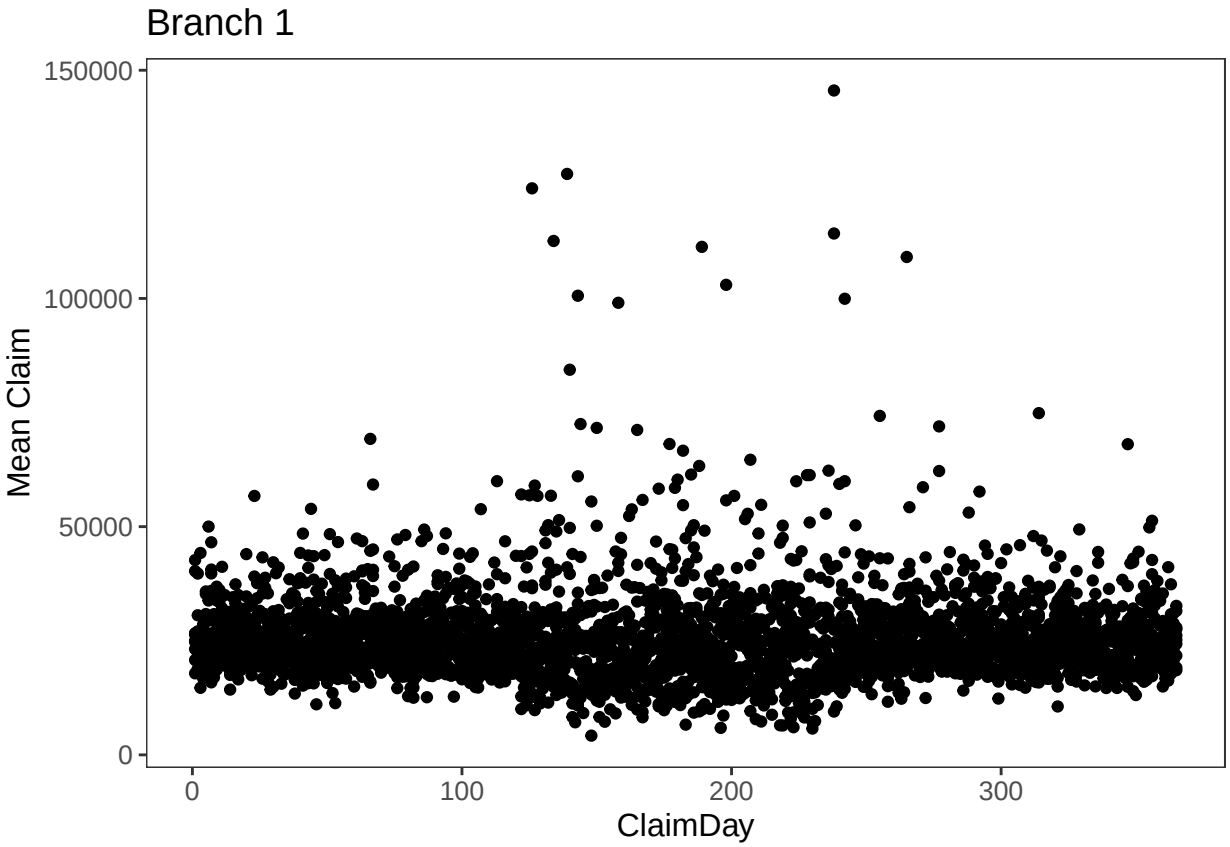
Exercise 2

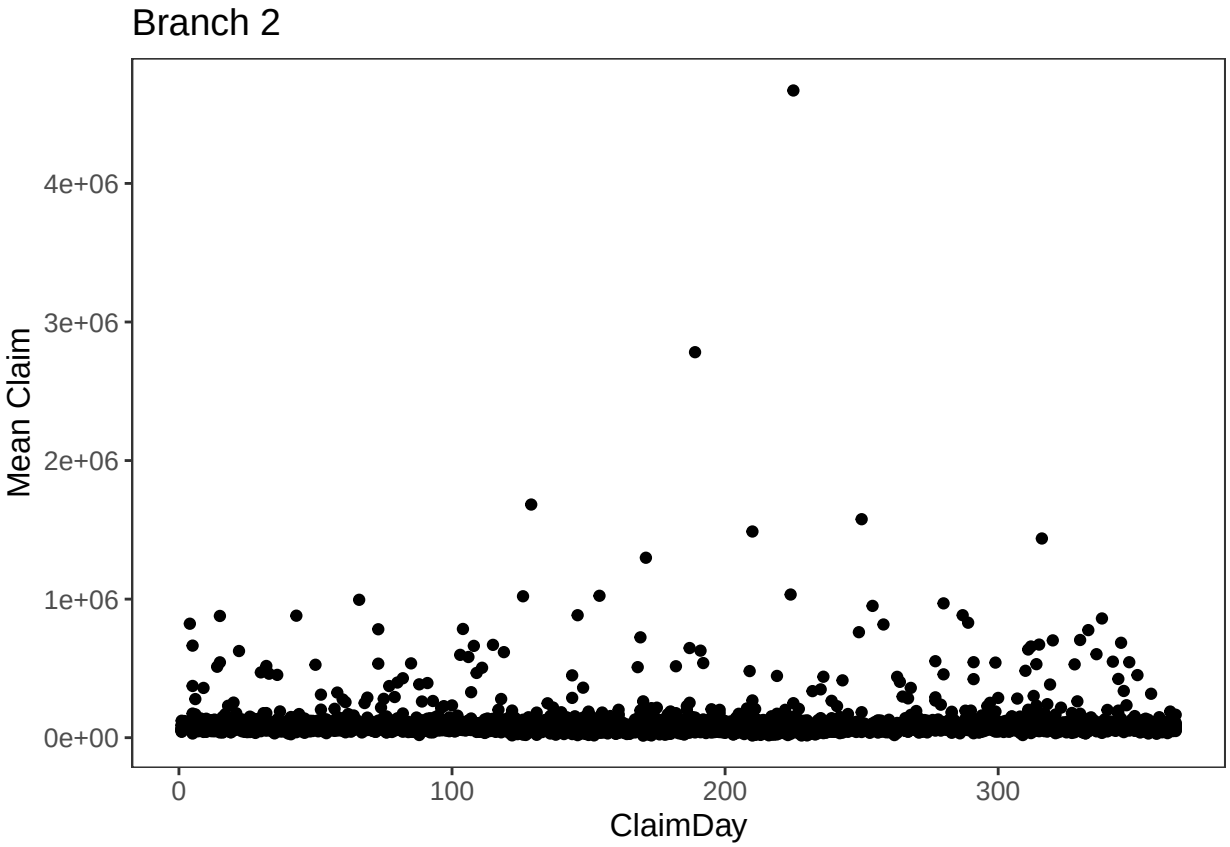
It is not only of interset to observe the number of claims over time but to also examine the sizes of these claims.


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

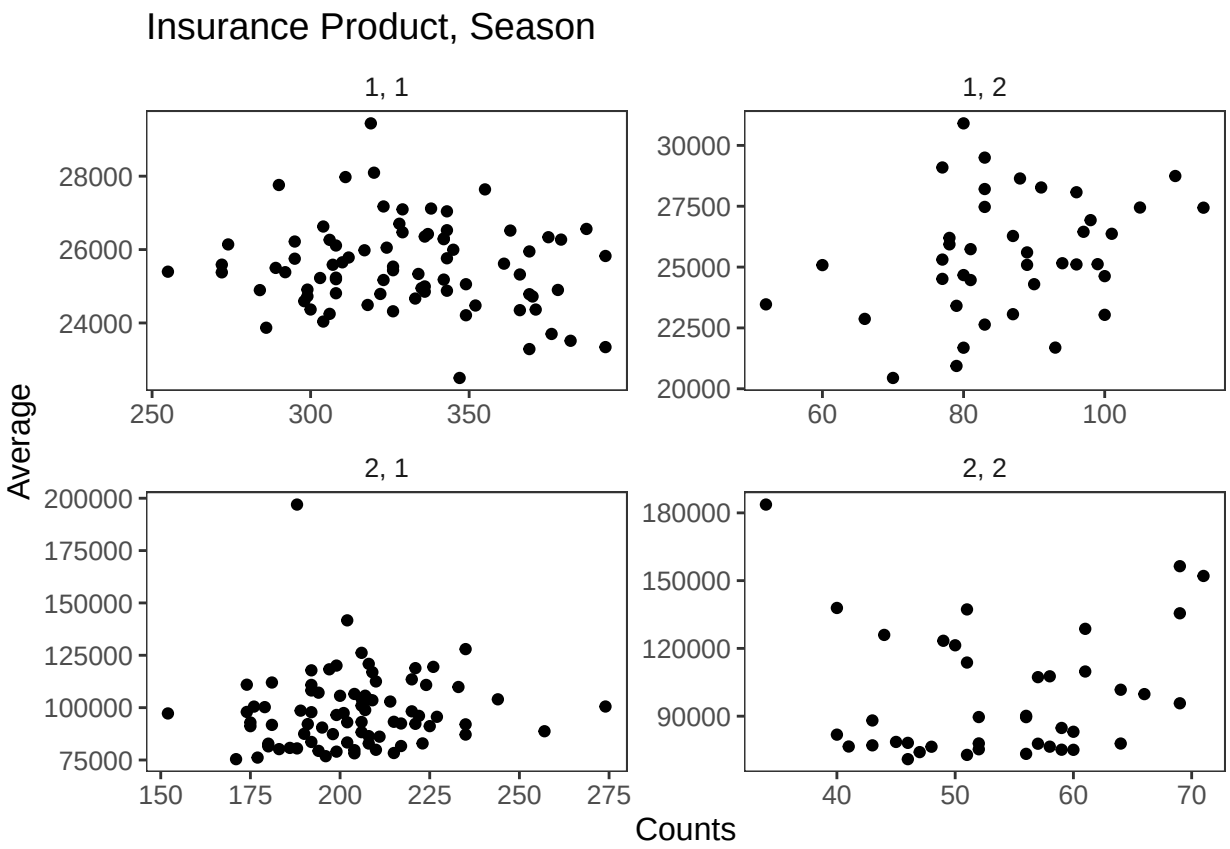


It seems as if the avergae claim costs are time independent, atleast when aggregated on a monthly basis, from the two previous plots. We can also observe the mean of the claim sizes for every day in order to further strengthen the assumption of time independence.


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

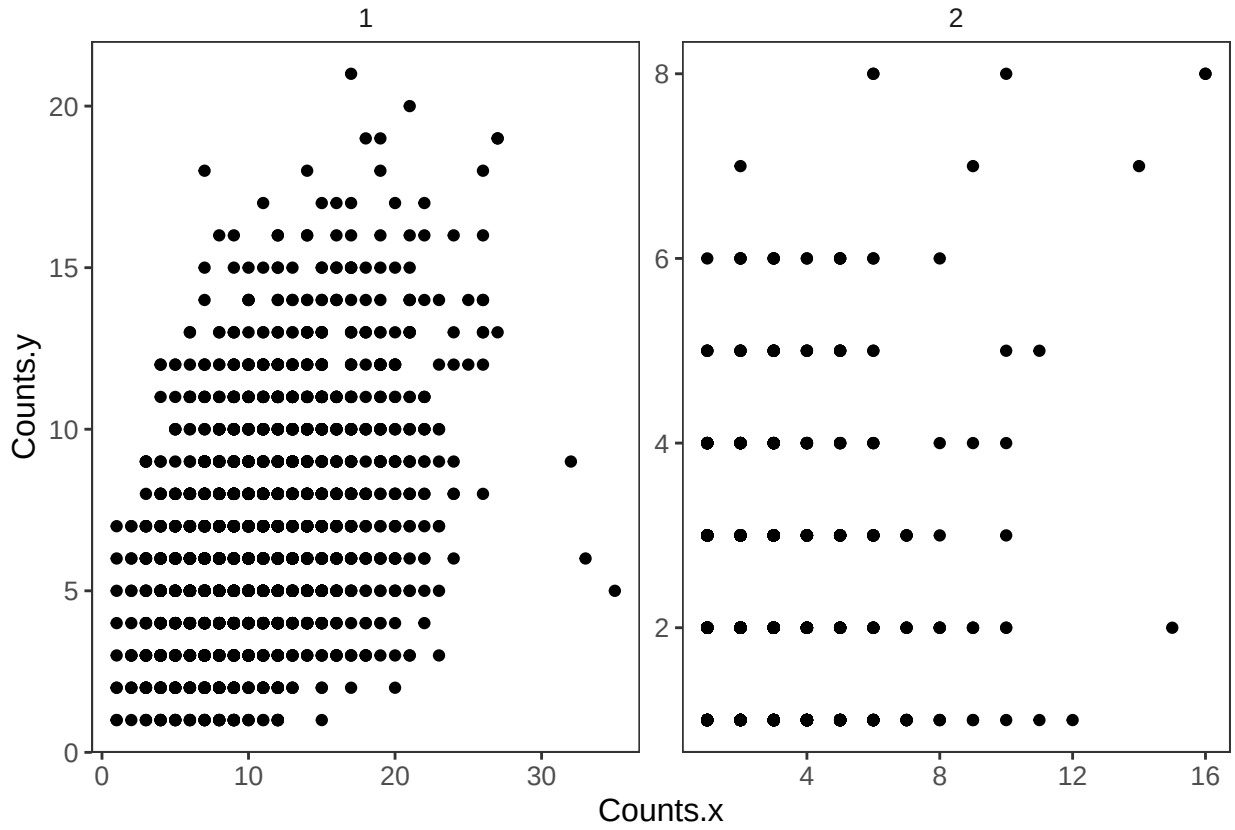


We note from the figures that the average claim sizes seem to be time independet. An interesting factor is however that all extremely deviating claim costs seem to occur during the summer (i.e period 2). In conclussion it seems as if the claim costs are independent of time. Another thing that we want 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.

A single simulation gives us the following number of claims to the coming year.

	n1	n2
1	295	179
2	304	175
3	272	176
4	336	197
5	88	34
6	52	57

	n1	n2
7	96	46
8	110	59
51	308	206
61	319	194
71	375	220
81	324	186