

Oblig 2, STK4100

Hanne Rønning Berg

2024-10-26

Problem 1

```
data="http://www.uio.no/studier/emner/matnat/math/STK3100/data/claims.txt"  
claims=read.table(data,header=T)
```

a)

Count data (antskader in this case) is usually modeled by a poisson distribution.

Some assumptions for the poisson distribution is:

The events can be counted. Antskader is integers.

The events are independent.

Constant rate

Also a poisson distribution have same mean and variance.

4 age categories, 4 motorvolume categories, 4 districts categories. $4 \times 4 \times 4 = 64$, so 64 combinations

The sum of poisson distributions observations is also poisson distributed.

b)

```
fit.pois.offset <- glm(antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(m  
otorvolum) + as.factor(distrikt),  
  family = poisson(link = "log"),  
  data = claims)  
  
summary(fit.pois.offset)
```

```
##
## Call:
## glm(formula = antskader ~ offset(log(antforsikret)) + as.factor(alder) +
##      as.factor(motorvolum) + as.factor(distrikt), family = poisson(link = "log"),
##      data = claims)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.82174     0.07679  -23.724 < 2e-16 ***
## as.factor(alder)2    -0.19101     0.08286   -2.305 0.021149 *
## as.factor(alder)3    -0.34495     0.08137   -4.239 2.24e-05 ***
## as.factor(alder)4    -0.53667     0.06996   -7.672 1.70e-14 ***
## as.factor(motorvolum)2  0.16134     0.05053    3.193 0.001409 **
## as.factor(motorvolum)3  0.39281     0.05500    7.142 9.18e-13 ***
## as.factor(motorvolum)4  0.56341     0.07232    7.791 6.65e-15 ***
## as.factor(distrikt)2   0.02587     0.04302    0.601 0.547597
## as.factor(distrikt)3   0.03852     0.05051    0.763 0.445657
## as.factor(distrikt)4   0.23421     0.06167    3.798 0.000146 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 236.26  on 63  degrees of freedom
## Residual deviance:  51.42  on 54  degrees of freedom
## AIC: 388.74
##
## Number of Fisher Scoring iterations: 4
```

Some of the 64 groups have more insurance takers than others.

We want to find claims per insured car, and the first thing that comes to mind would be to divide antskader by antforsikret. But, then we wouldn't have integers (counts) anymore, making a poisson model difficult.

In a poisson GLM: $\log(\mu/\text{antforsikret}) = \text{sum of betas} * x\text{'s}$ And that is the same as $\log(\mu) = \log(\text{antforsikret}) + \text{sum of betas} * x\text{'s}$.

So, we use an offset when our μ 's are proportional to some constant.

$\log(\text{antforsikret})$ enters the model with a fixed coefficient = 1.

c)

```
fit.pois.nooffset = glm(antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(distrikt), family = poisson(link = "log"), data = claims)
summary(fit.pois.nooffset)
```

```
##
## Call:
## glm(formula = antskader ~ as.factor(alder) + as.factor(motorvolum) +
##      as.factor(distrikt), family = poisson(link = "log"), data = claims)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.84305    0.07944  35.787 < 2e-16 ***
## as.factor(alder)2    0.56769    0.08272   6.863 6.74e-12 ***
## as.factor(alder)3    0.68217    0.08108   8.413 < 2e-16 ***
## as.factor(alder)4    2.19916    0.06965  31.575 < 2e-16 ***
## as.factor(motorvolum)2 0.98960    0.05045  19.617 < 2e-16 ***
## as.factor(motorvolum)3 0.47070    0.05490   8.574 < 2e-16 ***
## as.factor(motorvolum)4 -0.58927    0.07211  -8.172 3.04e-16 ***
## as.factor(distrikt)2 -0.43822    0.04297 -10.198 < 2e-16 ***
## as.factor(distrikt)3 -0.91521    0.05032 -18.187 < 2e-16 ***
## as.factor(distrikt)4 -1.44367    0.06158 -23.445 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 4236.68  on 63  degrees of freedom
## Residual deviance:  121.31  on 54  degrees of freedom
## AIC: 458.63
##
## Number of Fisher Scoring iterations: 4
```

The deviance is a lot higher compared to the model with an offset. A high residual deviance indicate overdispersion,

The AIC is also higher.

In this model all coefficients have a low p-value.

We will test the different covariates to see how significant they are:

```
#age:

fit.offset.noage = glm(antskader ~ offset(log(antforsikret)) + as.factor(motorvolum) + as.factor(distrikt),
                      family = poisson(link = "log"),
                      data = claims)

anova(fit.pois.offset, fit.offset.noage ,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(motorvolum)
+
##   as.factor(distrikt)
## Model 2: antskader ~ offset(log(antforsikret)) + as.factor(motorvolum) +
##   as.factor(distrikt)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         54      51.42
## 2         57     136.29 -3   -84.87 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#engine volume:

```
fit.offset.noengine = glm(antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(distrikt),
                          family = poisson(link = "log"),
                          data = claims)

anova(fit.pois.offset, fit.offset.noengine ,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(motorvolum)
+
##   as.factor(distrikt)
## Model 2: antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(distrikt)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         54      51.42
## 2         57     140.09 -3   -88.667 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#district:

```
fit.offset.nodistrict = glm(antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(motorvolum),
                            family = poisson(link = "log"),
                            data = claims)

anova(fit.pois.offset, fit.offset.nodistrict ,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(motorvolum)
+
##      as.factor(distrikt)
## Model 2: antskader ~ offset(log(antforsikret)) + as.factor(alder) + as.factor(motorvolum)
##      Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1          54      51.420
## 2          57      65.291 -3  -13.871 0.003086 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the model with an offset: Age, motor volume and district are all significant, even though district is a bit less.

```
fit.interaction = glm(antskader ~ offset(log(antforsikret)) * as.factor(alder) * as.factor(mo
torvolum) * as.factor(distrikt),
                      family = poisson(link = "log"),
                      data = claims)

summary(fit.interaction)
```

```
##
## Call:
## glm(formula = antskader ~ offset(log(antforsikret)) * as.factor(alder) *
##      as.factor(motorvolum) * as.factor(distrikt), family = poisson(link = "log"),
##      data = claims)
##
## Coefficients:
##
##                                     Estimate
## (Intercept)                        -1.646e+00
## as.factor(alder)2                  -3.750e-01
## as.factor(alder)3                  -8.640e-01
## as.factor(alder)4                  -7.311e-01
## as.factor(motorvolum)2              1.398e-01
## as.factor(motorvolum)3             -3.003e-01
## as.factor(motorvolum)4             -1.461e-01
## as.factor(distrikt)2                2.940e-01
## as.factor(distrikt)3               -3.003e-01
## as.factor(distrikt)4               -6.570e-01
## as.factor(alder)2:as.factor(motorvolum)2  2.751e-02
## as.factor(alder)3:as.factor(motorvolum)2  3.131e-01
## as.factor(alder)4:as.factor(motorvolum)2  4.470e-02
## as.factor(alder)2:as.factor(motorvolum)3  6.161e-01
## as.factor(alder)3:as.factor(motorvolum)3  1.242e+00
## as.factor(alder)4:as.factor(motorvolum)3  7.256e-01
## as.factor(alder)2:as.factor(motorvolum)4  7.944e-01
## as.factor(alder)3:as.factor(motorvolum)4  1.005e+00
## as.factor(alder)4:as.factor(motorvolum)4  7.530e-01
## as.factor(alder)2:as.factor(distrikt)2   -2.634e-01
## as.factor(alder)3:as.factor(distrikt)2    2.893e-01
## as.factor(alder)4:as.factor(distrikt)2   -2.877e-01
## as.factor(alder)2:as.factor(distrikt)3    4.283e-01
## as.factor(alder)3:as.factor(distrikt)3    6.238e-01
## as.factor(alder)4:as.factor(distrikt)3    4.078e-01
## as.factor(alder)2:as.factor(distrikt)4    7.905e-01
## as.factor(alder)3:as.factor(distrikt)4    8.640e-01
## as.factor(alder)4:as.factor(distrikt)4    8.614e-01
## as.factor(motorvolum)2:as.factor(distrikt)2 -5.732e-01
## as.factor(motorvolum)3:as.factor(distrikt)2  1.013e-01
## as.factor(motorvolum)4:as.factor(distrikt)2  6.868e-01
## as.factor(motorvolum)2:as.factor(distrikt)3  1.384e-01
## as.factor(motorvolum)3:as.factor(distrikt)3  1.148e+00
## as.factor(motorvolum)4:as.factor(distrikt)3  1.245e+00
## as.factor(motorvolum)2:as.factor(distrikt)4  6.747e-01
## as.factor(motorvolum)3:as.factor(distrikt)4  1.322e+00
## as.factor(motorvolum)4:as.factor(distrikt)4 -2.095e+01
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)2  5.816e-01
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)2 -9.946e-02
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)2  6.280e-01
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)2  2.367e-01
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)2 -8.512e-01
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)2 -2.055e-01
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)2 -5.082e-01
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)2 -1.411e+00
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)2 -7.276e-01
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)3 -2.785e-01
```

```

## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)3 -2.750e-01
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)3 -2.220e-01
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)3 -9.831e-01
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)3 -1.521e+00
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)3 -1.228e+00
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)3 -2.675e+00
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)3 -1.599e+00
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)3 -1.473e+00
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)4 -1.047e+00
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)4 -5.380e-01
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)4 -6.468e-01
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)4 -1.468e+00
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)4 -1.408e+00
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)4 -1.273e+00
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)4 2.121e+01
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)4 2.126e+01
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)4 2.128e+01
## Std. Error
## (Intercept) 1.622e-01
## as.factor(alder)2 2.343e-01
## as.factor(alder)3 2.762e-01
## as.factor(alder)4 1.809e-01
## as.factor(motorvolum)2 2.054e-01
## as.factor(motorvolum)3 2.810e-01
## as.factor(motorvolum)4 5.257e-01
## as.factor(distrikt)2 2.679e-01
## as.factor(distrikt)3 4.757e-01
## as.factor(distrikt)4 7.255e-01
## as.factor(alder)2:as.factor(motorvolum)2 2.875e-01
## as.factor(alder)3:as.factor(motorvolum)2 3.216e-01
## as.factor(alder)4:as.factor(motorvolum)2 2.261e-01
## as.factor(alder)2:as.factor(motorvolum)3 3.560e-01
## as.factor(alder)3:as.factor(motorvolum)3 3.774e-01
## as.factor(alder)4:as.factor(motorvolum)3 2.994e-01
## as.factor(alder)2:as.factor(motorvolum)4 6.004e-01
## as.factor(alder)3:as.factor(motorvolum)4 6.156e-01
## as.factor(alder)4:as.factor(motorvolum)4 5.438e-01
## as.factor(alder)2:as.factor(distrikt)2 3.911e-01
## as.factor(alder)3:as.factor(distrikt)2 4.089e-01
## as.factor(alder)4:as.factor(distrikt)2 2.995e-01
## as.factor(alder)2:as.factor(distrikt)3 5.880e-01
## as.factor(alder)3:as.factor(distrikt)3 6.135e-01
## as.factor(alder)4:as.factor(distrikt)3 4.976e-01
## as.factor(alder)2:as.factor(distrikt)4 8.688e-01
## as.factor(alder)3:as.factor(distrikt)4 9.090e-01
## as.factor(alder)4:as.factor(distrikt)4 7.487e-01
## as.factor(motorvolum)2:as.factor(distrikt)2 3.573e-01
## as.factor(motorvolum)3:as.factor(distrikt)2 4.425e-01
## as.factor(motorvolum)4:as.factor(distrikt)2 7.561e-01
## as.factor(motorvolum)2:as.factor(distrikt)3 5.850e-01
## as.factor(motorvolum)3:as.factor(distrikt)3 6.356e-01
## as.factor(motorvolum)4:as.factor(distrikt)3 8.998e-01
## as.factor(motorvolum)2:as.factor(distrikt)4 8.277e-01
## as.factor(motorvolum)3:as.factor(distrikt)4 8.826e-01
## as.factor(motorvolum)4:as.factor(distrikt)4 4.225e+04
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)2 4.903e-01

```

```

## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)2 5.047e-01
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)2 3.892e-01
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)2 5.639e-01
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)2 5.748e-01
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)2 4.744e-01
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)2 8.804e-01
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)2 8.962e-01
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)2 7.884e-01
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)3 7.178e-01
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)3 7.283e-01
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)3 6.094e-01
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)3 7.715e-01
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)3 7.805e-01
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)3 6.629e-01
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)3 1.218e+00
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)3 1.066e+00
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)3 9.333e-01
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)4 1.013e+00
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)4 1.021e+00
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)4 8.553e-01
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)4 1.081e+00
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)4 1.075e+00
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)4 9.129e-01
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)4 4.225e+04
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)4 4.225e+04
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)4 4.225e+04
##
## (Intercept) -10.144 < 2e-16
## as.factor(alder)2 -1.601 0.10947
## as.factor(alder)3 -3.128 0.00176
## as.factor(alder)4 -4.041 5.32e-05
## as.factor(motorvolum)2 0.681 0.49618
## as.factor(motorvolum)3 -1.069 0.28518
## as.factor(motorvolum)4 -0.278 0.78100
## as.factor(distrikt)2 1.097 0.27244
## as.factor(distrikt)3 -0.631 0.52789
## as.factor(distrikt)4 -0.906 0.36516
## as.factor(alder)2:as.factor(motorvolum)2 0.096 0.92379
## as.factor(alder)3:as.factor(motorvolum)2 0.974 0.33025
## as.factor(alder)4:as.factor(motorvolum)2 0.198 0.84324
## as.factor(alder)2:as.factor(motorvolum)3 1.731 0.08351
## as.factor(alder)3:as.factor(motorvolum)3 3.290 0.00100
## as.factor(alder)4:as.factor(motorvolum)3 2.423 0.01538
## as.factor(alder)2:as.factor(motorvolum)4 1.323 0.18575
## as.factor(alder)3:as.factor(motorvolum)4 1.633 0.10254
## as.factor(alder)4:as.factor(motorvolum)4 1.385 0.16616
## as.factor(alder)2:as.factor(distrikt)2 -0.674 0.50059
## as.factor(alder)3:as.factor(distrikt)2 0.708 0.47920
## as.factor(alder)4:as.factor(distrikt)2 -0.961 0.33674
## as.factor(alder)2:as.factor(distrikt)3 0.728 0.46637
## as.factor(alder)3:as.factor(distrikt)3 1.017 0.30918
## as.factor(alder)4:as.factor(distrikt)3 0.819 0.41254
## as.factor(alder)2:as.factor(distrikt)4 0.910 0.36291
## as.factor(alder)3:as.factor(distrikt)4 0.950 0.34188
## as.factor(alder)4:as.factor(distrikt)4 1.151 0.24989
## as.factor(motorvolum)2:as.factor(distrikt)2 -1.604 0.10861

```



```

## as.factor(motorvolum)3:as.factor(distrikt)2      0.229  0.81893
## as.factor(motorvolum)4:as.factor(distrikt)2      0.908  0.36372
## as.factor(motorvolum)2:as.factor(distrikt)3      0.237  0.81294
## as.factor(motorvolum)3:as.factor(distrikt)3      1.806  0.07098
## as.factor(motorvolum)4:as.factor(distrikt)3      1.383  0.16655
## as.factor(motorvolum)2:as.factor(distrikt)4      0.815  0.41495
## as.factor(motorvolum)3:as.factor(distrikt)4      1.498  0.13418
## as.factor(motorvolum)4:as.factor(distrikt)4      0.000  0.99960
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)2  1.186  0.23549
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)2 -0.197  0.84378
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)2  1.613  0.10664
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)2  0.420  0.67463
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)2 -1.481  0.13866
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)2 -0.433  0.66485
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)2 -0.577  0.56378
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)2 -1.575  0.11533
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)2 -0.923  0.35609
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)3 -0.388  0.69801
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)3 -0.378  0.70576
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)3 -0.364  0.71564
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)3 -1.274  0.20257
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)3 -1.949  0.05134
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)3 -1.853  0.06394
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)3 -2.195  0.02816
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)3 -1.500  0.13368
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)3 -1.578  0.11456
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)4 -1.034  0.30128
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)4 -0.527  0.59811
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)4 -0.756  0.44951
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)4 -1.358  0.17454
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)4 -1.310  0.19020
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)4 -1.394  0.16332
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)4  0.001  0.99960
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)4  0.001  0.99960
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)4  0.001  0.99960
##
## (Intercept)                                     ***
## as.factor(alder)2
## as.factor(alder)3                               **
## as.factor(alder)4                               ***
## as.factor(motorvolum)2
## as.factor(motorvolum)3
## as.factor(motorvolum)4
## as.factor(distrikt)2
## as.factor(distrikt)3
## as.factor(distrikt)4
## as.factor(alder)2:as.factor(motorvolum)2
## as.factor(alder)3:as.factor(motorvolum)2
## as.factor(alder)4:as.factor(motorvolum)2
## as.factor(alder)2:as.factor(motorvolum)3        .
## as.factor(alder)3:as.factor(motorvolum)3        **
## as.factor(alder)4:as.factor(motorvolum)3        *
## as.factor(alder)2:as.factor(motorvolum)4
## as.factor(alder)3:as.factor(motorvolum)4
## as.factor(alder)4:as.factor(motorvolum)4
## as.factor(alder)2:as.factor(distrikt)2

```

```

## as.factor(alder)3:as.factor(distrikt)2
## as.factor(alder)4:as.factor(distrikt)2
## as.factor(alder)2:as.factor(distrikt)3
## as.factor(alder)3:as.factor(distrikt)3
## as.factor(alder)4:as.factor(distrikt)3
## as.factor(alder)2:as.factor(distrikt)4
## as.factor(alder)3:as.factor(distrikt)4
## as.factor(alder)4:as.factor(distrikt)4
## as.factor(motorvolum)2:as.factor(distrikt)2
## as.factor(motorvolum)3:as.factor(distrikt)2
## as.factor(motorvolum)4:as.factor(distrikt)2
## as.factor(motorvolum)2:as.factor(distrikt)3
## as.factor(motorvolum)3:as.factor(distrikt)3
## as.factor(motorvolum)4:as.factor(distrikt)3
## as.factor(motorvolum)2:as.factor(distrikt)4
## as.factor(motorvolum)3:as.factor(distrikt)4
## as.factor(motorvolum)4:as.factor(distrikt)4
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)2
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)2
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)2
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)2
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)2
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)2
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)2
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)2
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)2
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)3
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)3
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)3
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)3
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)3
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)3
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)3 *
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)3
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)3
## as.factor(alder)2:as.factor(motorvolum)2:as.factor(distrikt)4
## as.factor(alder)3:as.factor(motorvolum)2:as.factor(distrikt)4
## as.factor(alder)4:as.factor(motorvolum)2:as.factor(distrikt)4
## as.factor(alder)2:as.factor(motorvolum)3:as.factor(distrikt)4
## as.factor(alder)3:as.factor(motorvolum)3:as.factor(distrikt)4
## as.factor(alder)4:as.factor(motorvolum)3:as.factor(distrikt)4
## as.factor(alder)2:as.factor(motorvolum)4:as.factor(distrikt)4
## as.factor(alder)3:as.factor(motorvolum)4:as.factor(distrikt)4
## as.factor(alder)4:as.factor(motorvolum)4:as.factor(distrikt)4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2.3626e+02  on 63  degrees of freedom
## Residual deviance: 4.1233e-10  on  0  degrees of freedom
## AIC: 445.32
##
## Number of Fisher Scoring iterations: 20

```

This is a model checking for all possible interaction terms. Age group 3 and 4 seems to have a significant value when interacted with motor volume 3. Also age 2 and motor volume 4.

```
#Using model selection on the model with interaction terms.
stepwise_model <- step(fit.interaction, direction = "back")
```

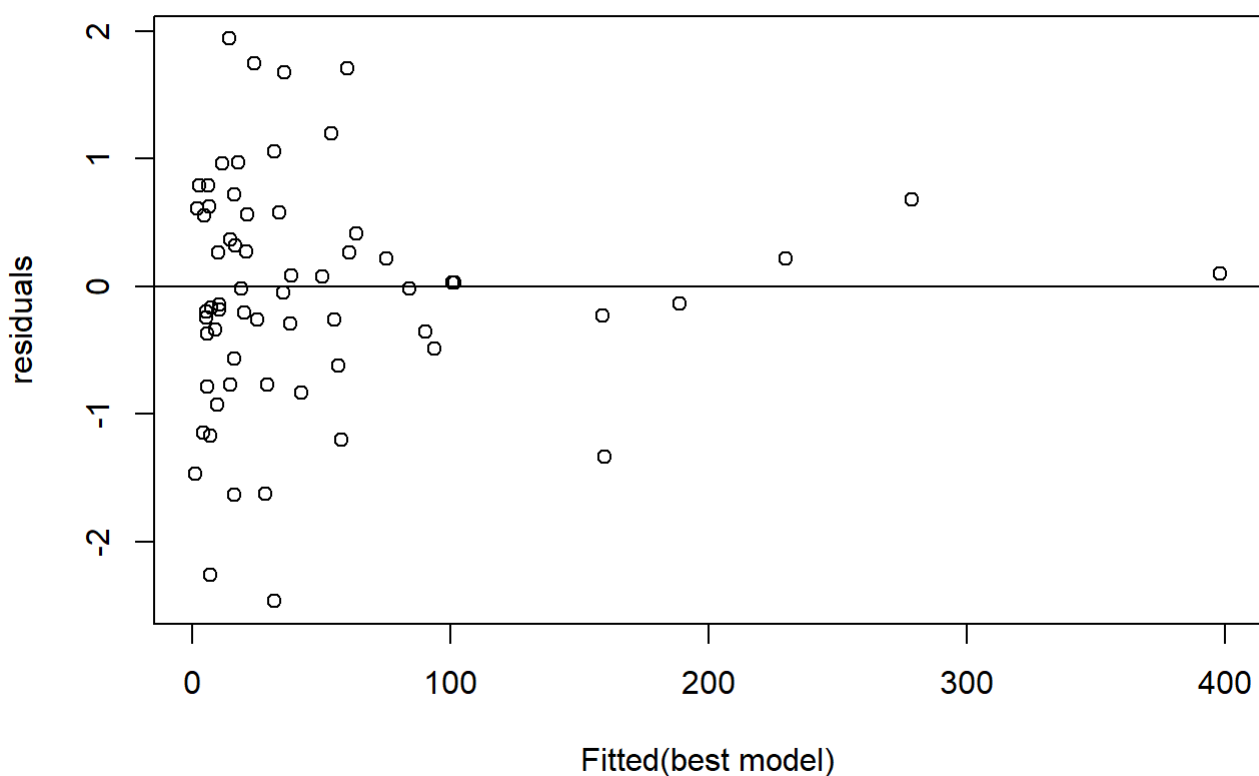
```
## Start: AIC=445.32
## antskader ~ offset(log(antforsikret)) * as.factor(alder) * as.factor(motorvolum) *
##   as.factor(distrikt)
##
##                                     Df Deviance    AIC
## - as.factor(alder):as.factor(motorvolum):as.factor(distrikt) 27    27.29 418.61
## <none>                                                         0.00 445.32
##
## Step: AIC=418.61
## antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(distrikt) +
##   as.factor(alder):as.factor(motorvolum) + as.factor(alder):as.factor(distrikt) +
##   as.factor(motorvolum):as.factor(distrikt) + offset(log(antforsikret))
##
##                                     Df Deviance    AIC
## - as.factor(alder):as.factor(distrikt)      9    33.527 406.85
## - as.factor(motorvolum):as.factor(distrikt)  9    34.457 407.78
## - as.factor(alder):as.factor(motorvolum)     9    37.685 411.01
## <none>                                       27.290 418.61
##
## Step: AIC=406.85
## antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(distrikt) +
##   as.factor(alder):as.factor(motorvolum) + as.factor(motorvolum):as.factor(distrikt) +
##   offset(log(antforsikret))
##
##                                     Df Deviance    AIC
## - as.factor(motorvolum):as.factor(distrikt)  9    40.907 396.23
## - as.factor(alder):as.factor(motorvolum)     9    44.132 399.45
## <none>                                       33.527 406.85
##
## Step: AIC=396.23
## antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(distrikt) +
##   as.factor(alder):as.factor(motorvolum) + offset(log(antforsikret))
##
##                                     Df Deviance    AIC
## - as.factor(alder):as.factor(motorvolum)     9    51.420 388.74
## <none>                                       40.907 396.23
## - as.factor(distrikt)                       3    54.850 404.17
##
## Step: AIC=388.74
## antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(distrikt) +
##   offset(log(antforsikret))
##
##                                     Df Deviance    AIC
## <none>                                       51.420 388.74
## - as.factor(distrikt)      3    65.291 396.61
## - as.factor(alder)         3   136.290 467.61
## - as.factor(motorvolum)    3   140.087 471.41
```

The model with the lowest AIC is the model without any interaction terms. So the model with an offset, all covariates and no interaction terms seems to give the best description of the data.

d)

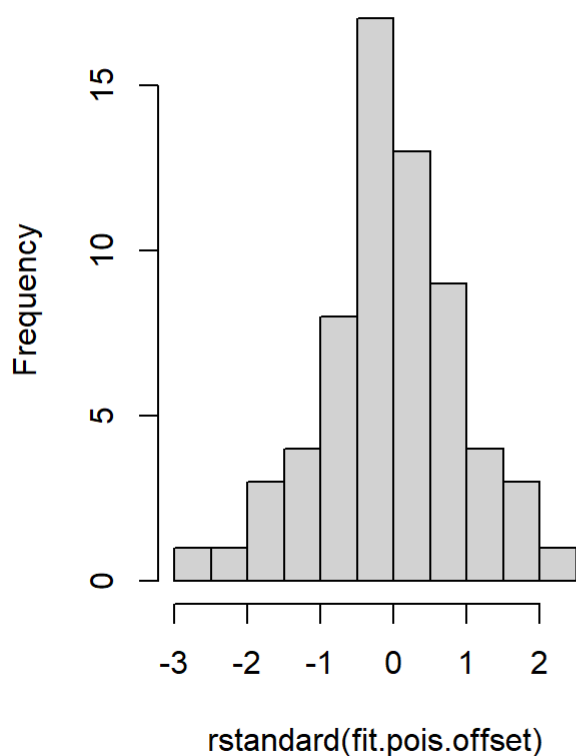
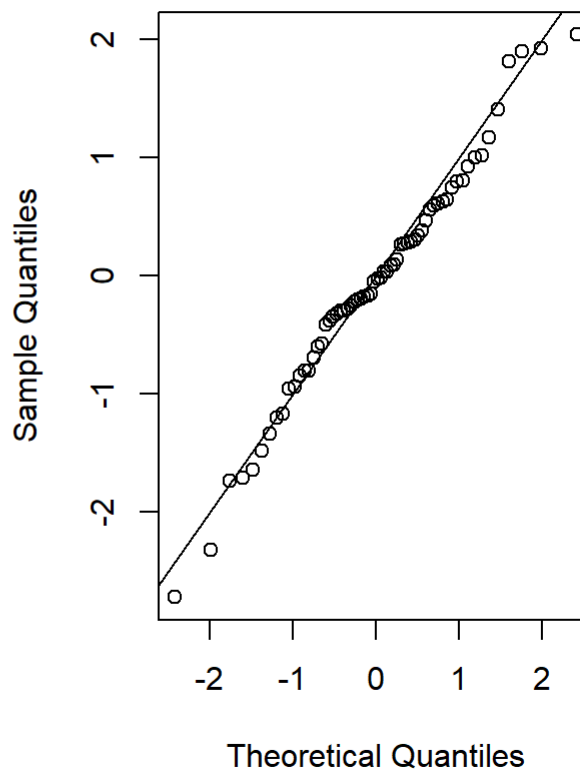
```
#produce residual vs. fitted plot
plot(fitted(fit.pois.offset), resid(fit.pois.offset), ylab="residuals", xlab="Fitted(best model)")

#add a horizontal line at 0
abline(0,0)
```



We want the residual values to be randomly around the horizontal axis. Any patterns in the plot, could indicate that a different model would be more suitable. The data points are mostly from 0 to 100, and that's because most instances in our data set is from 0-100.

```
par(mfrow = c(1, 2))
hist(rstandard(fit.pois.offset))
qqnorm(rstandard(fit.pois.offset))
abline(0, 1)
```

Histogram of rstandard(fit.pois.offsets)**Normal Q-Q Plot**

Histogram and Q-Q plot of standardized residuals.

The standardized residuals should have an approximately $N(0,1)$ distribution when the model holds (note that with fewer data points one should be careful about interpreting this as the residuals are affected by ordinary sampling variability).

For the Q-Q plot we want the data to form a straight line, as it does in the plot. This is an indication that we have a good model.

e)

Interpret the estimates from the best model in c as rate ratios, and give 95% cond intervalls for the ratios.

```
#from GPT:

coef_summary = summary(fit.pois.offset)$coefficients

# Compute the rate ratios and 95% confidence intervals
rate_ratios <- exp(coef_summary[, "Estimate"])
conf_int_lower <- exp(coef_summary[, "Estimate"] - 1.96 * coef_summary[, "Std. Error"])
conf_int_upper <- exp(coef_summary[, "Estimate"] + 1.96 * coef_summary[, "Std. Error"])

# Combine the results into a data frame
results <- data.frame(
  Estimate = coef_summary[, "Estimate"],
  Std_Error = coef_summary[, "Std. Error"],
  Rate_Ratio = rate_ratios,
  CI_Lower = conf_int_lower,
  CI_Upper = conf_int_upper
)

# Print the results
print(results)
```

##	Estimate	Std_Error	Rate_Ratio	CI_Lower	CI_Upper
## (Intercept)	-1.82173992	0.07678762	0.1617441	0.1391443	0.1880145
## as.factor(alder)2	-0.19101011	0.08285644	0.8261242	0.7022901	0.9717939
## as.factor(alder)3	-0.34495066	0.08137413	0.7082553	0.6038412	0.8307243
## as.factor(alder)4	-0.53667071	0.06995562	0.5846916	0.5097761	0.6706166
## as.factor(motorvolum)2	0.16133698	0.05053239	1.1750809	1.0642747	1.2974236
## as.factor(motorvolum)3	0.39281049	0.05499780	1.4811377	1.3297817	1.6497210
## as.factor(motorvolum)4	0.56341234	0.07231533	1.7566566	1.5245119	2.0241511
## as.factor(distrikt)2	0.02586819	0.04301579	1.0262057	0.9432322	1.1164781
## as.factor(distrikt)3	0.03852393	0.05051157	1.0392756	0.9413138	1.1474322
## as.factor(distrikt)4	0.23420533	0.06167328	1.2639040	1.1199967	1.4263018

f)

```
new.covar=data.frame(alder = 2, motorvolum = 2, distrikt = 1, antforsikret = 1)
prediction = predict(fit.pois.offset, newdata=new.covar, type="link", se.fit=TRUE)

rate_log = prediction$fit
se_log = prediction$se.fit

ci_lower_log = rate_log - 1.96 * se_log
ci_upper_log = rate_log + 1.96 * se_log

#exponentiate to transform to rate

cat(exp(rate_log), exp(ci_lower_log), exp(ci_upper_log))
```

```
## 0.1570151 0.1404317 0.1755569
```

The expected rate claim for the person is 0.157

And the 95% confidence interval for the person is [0.140, 0.176]

Problem 2

Solutions to task a) to e) is given as handwritten notes. See end of the submission.

f)

```
library(MASS)
```

```
fit.negbin = glm.nb(antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(distrikt) + offset(log(antforsikret)), data = claims)
```

```
## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, trace =  
## control$trace > : iteration limit reached  
## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, trace =  
## control$trace > : iteration limit reached
```

```
summary(fit.negbin)
```

```
##
## Call:
## glm.nb(formula = antskader ~ as.factor(alder) + as.factor(motorvolum) +
##       as.factor(distrikt) + offset(log(antforsikret)), data = claims,
##       init.theta = 449933.2767, link = log)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.82174    0.07679 -23.723  < 2e-16 ***
## as.factor(alder)2   -0.19101    0.08286  -2.305 0.021155 *
## as.factor(alder)3   -0.34495    0.08138  -4.239 2.25e-05 ***
## as.factor(alder)4   -0.53667    0.06996  -7.671 1.70e-14 ***
## as.factor(motorvolum)2  0.16133    0.05054   3.192 0.001412 **
## as.factor(motorvolum)3  0.39281    0.05500   7.142 9.23e-13 ***
## as.factor(motorvolum)4  0.56341    0.07232   7.791 6.67e-15 ***
## as.factor(distrikt)2   0.02587    0.04302   0.601 0.547637
## as.factor(distrikt)3   0.03853    0.05052   0.763 0.445686
## as.factor(distrikt)4   0.23421    0.06168   3.797 0.000146 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(449933.3) family taken to be 1)
##
##      Null deviance: 236.212  on 63  degrees of freedom
## Residual deviance:  51.416  on 54  degrees of freedom
## AIC: 390.74
##
## Number of Fisher Scoring iterations: 1
##
##              Theta: 449933
##             Std. Err.: 4185442
## Warning while fitting theta: iteration limit reached
##
## 2 x log-likelihood: -368.745
```

The AIC is 390.74 which is higher than the AIC from the best model in problem 1 (AIC: 388.74)

What does the estimated k (theta) tell you about possible over-dispersion.

Theta: 449933, the lower the theta, the higher the over-dispersion is. This is a higher theta, indication low over-dispersion. In this model we have an offset, that might take into account for the dispersion.

Like in problem e), when $k \rightarrow \infty$, the data is poisson distributed.

```
fit.negbin.nooffset = glm.nb(antskader ~ as.factor(alder) + as.factor(motorvolum) + as.factor(
distrikt), data = claims)
```

```
## Warning in glm.nb(antskader ~ as.factor(alder) + as.factor(motorvolum) + :
## alternation limit reached
```

```
summary(fit.negbin.nooffset)
```



```
##
## Call:
## glm.nb(formula = antskader ~ as.factor(alder) + as.factor(motorvolum) +
##       as.factor(distrikt), data = claims, init.theta = 48.80426084,
##       link = log)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.86307    0.10487  27.300 < 2e-16 ***
## as.factor(alder)2    0.58188    0.10330   5.633 1.77e-08 ***
## as.factor(alder)3    0.71123    0.10172   6.992 2.71e-12 ***
## as.factor(alder)4    2.24889    0.09196  24.455 < 2e-16 ***
## as.factor(motorvolum)2 0.94631    0.08082  11.709 < 2e-16 ***
## as.factor(motorvolum)3 0.47769    0.08418   5.675 1.39e-08 ***
## as.factor(motorvolum)4 -0.59477    0.09827  -6.052 1.43e-09 ***
## as.factor(distrikt)2 -0.46156    0.07492  -6.160 7.25e-10 ***
## as.factor(distrikt)3 -1.00673    0.08157 -12.342 < 2e-16 ***
## as.factor(distrikt)4 -1.50639    0.09018 -16.704 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(48.8043) family taken to be 1)
##
## Null deviance: 1946.065 on 63 degrees of freedom
## Residual deviance: 79.061 on 54 degrees of freedom
## AIC: 453.49
##
## Number of Fisher Scoring iterations: 1
##
##              Theta: 48.8
##             Std. Err.: 25.5
## Warning while fitting theta: alternation limit reached
##
## 2 x log-likelihood: -431.486
```

For a nb.glm without offset the theta is much lower, indication an over-dispersion. Here the AIC is 453.49. Higher than the other models.

The poisson GLM with an offset is a better model, than negative binomial distribution.