# Lecture 6: Poisson regression

Claudia Czado

TU München

# Overview

- Introduction
- EDA for Poisson regression
- Estimation and testing in Poisson regression
- Residuals in Poisson regression

# Introduction

Want regression models for count response data. See Cameron and Trivedi (1998) and Winkelmann (1997) for books on regression models for count data.

**Example:** Canadian car insurance

$Y_i =$ Number of claims in 1957-1958 within a group $i$ of insured with $t_i$ insurance years

$\mathbf{x_i} =$ covariate of group $i$

# Poisson regression

**Model:**

$$
\begin{aligned}
Y_i &\sim Poisson(t_i \mu(\mathbf{x_i}, \boldsymbol{\beta})) \quad i = 1, \ldots, n \quad ind., \qquad i.e. \\
P(Y_i = y_i) &= e^{-t_i \mu(\mathbf{x_i}, \beta)} \frac{(t_i \mu(\mathbf{x_i}, \beta))^{y_i}}{y_i!} \\
\mu(\mathbf{x_i}, \boldsymbol{\beta}) &:= e^{\mathbf{x_i}^t \beta} \geq 0 \\
E(Y_i) &= t_i \mu(\mathbf{x_i}, \boldsymbol{\beta}) = Var(Y_i)
\end{aligned}
$$

$t_i$ gives the time length in which events occur, $t_i$ known.

## Likelihood:

$$
l(\mathbf{y}, \boldsymbol{\beta}) = \prod_{i=1}^{n} P(Y_i = y_i) = e^{-\sum_{i=1}^{n} t_i \mu(\mathbf{x_i}, \beta)} \prod_{i=1}^{n} \frac{(t_i \mu(\mathbf{x_i}, \beta))^{y_i}}{y_i!}
$$

# EDA for Poisson regression

For the Poisson model

$$Y_i \sim Poisson(t_i e^{\mathbf{x_i}^t \beta}) \quad ind.$$

$$\Rightarrow \quad \ln(\mu_i) = \ln(t_i) + \boldsymbol{x_i^t \beta} \qquad (*)$$

If we only have a single obs. for each $\mathbf{x_i}$ we can estimate $(*)$ by

$$\ln(Y_i) = \ln(t_i) + \boldsymbol{x_i^t \beta}$$

Therefore $\ln(Y_i) - \ln(t_i)$ needs to be linear in $\mathbf{x_i}$ if the Poisson model is valid. If $Y_i = 0$ holds we need to consider $\ln(Y_i + c)$ for $c$ small. Similar as for the binary regression we can investigate main and interaction effects.

If there are several obs. denoted by $Y_{i1}, \dots, Y_{in_i}$ with $t_{i1}, \dots, t_{in_i}$ for each design vector $\mathbf{x_i}$ we can plot

$$\mathbf{x_i} \qquad \text{versus} \qquad \left[\ln\left(\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}\right) - \ln\left(\frac{1}{n_i}\sum_{j=1}^{n_i} t_{ij}\right)\right]$$

In this case we can check whether $E(Y_i) = Var(Y_i)$ is valid for Poisson regression:

$$\text{plot} \qquad \hat{\mu}_i^0 = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \qquad \text{versus} \qquad s_i^0 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i^0)^2$$

and we expect this plot to be centered around the $45^o$- line.

# Example: Canadian Automobile Insurance Claims

Source: Bailey and Simon (1960)

Description: The data give the Canadian automobile insurance experience for policy years 1956 and 1957 as of June 30, 1959. The data includes virtually every insurance company operating in Canada and was collated by the Statistical Agency (Canadian Underwriters' Association - Statistical Department) acting under instructions from the Superintendent of Insurance. The data given here is for private passenger automobile liability for non-farmers for all of Canada excluding Saskatchewan.

The variable Merit measures the number of years since the last claim on the policy. The variable Class is a collation of age, sex, use and marital status. The variables Insured and Premium are two measures of the risk exposure of the insurance companies.

# Variable Description:

Merit

| | |
|---|---|
| 3 | licensed and accident free $\geq$3 years |
| 2 | licensed and accident free 2 years |
| 1 | licensed and accident free 1 year |
| 0 | all others |

Class

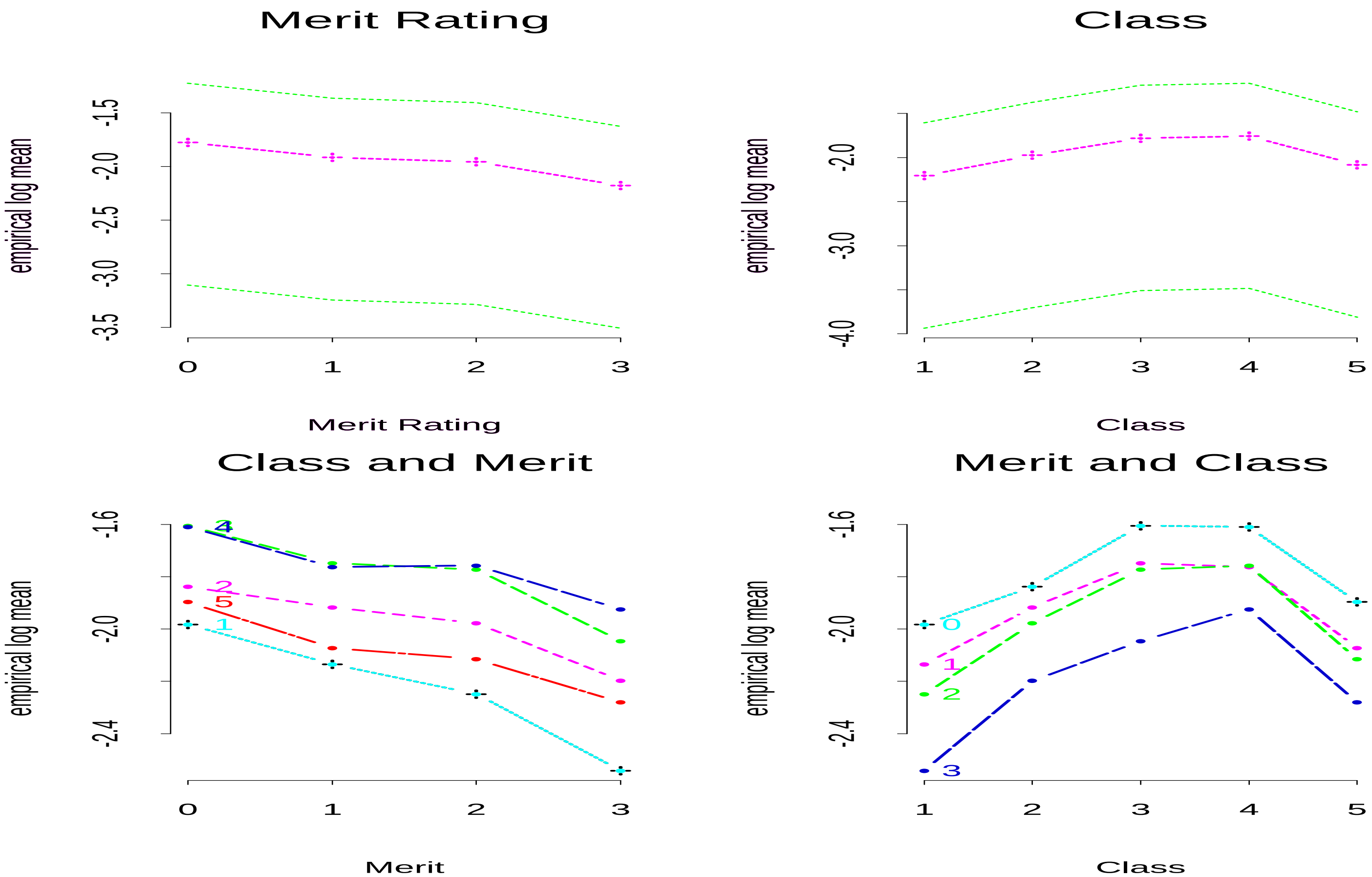
| | |
|---|---|
| 1 | pleasure, no male operator $< 25$ |
| 2 | pleasure, non-principal male operator $< 25$ |
| 3 | business use |
| 4 | unmarried owner and principal operator$< 25$ |
| 5 | married owner and principal operator $< 25$ |

| | |
|---|---|
| Insured | Earned car years |
| Premium | Earned premium in 1000's (adjusted to what the premium would have been had all cars been written at 01 rates) |
| Claims | Number of claims |
| Cost | Total cost of the claim in 1000's of dollars |

# Data

```
> cancar.data
   Merit Class Insured Premium Claims  Cost
 1     3     1 2757520  159108 217151 63191
 2     3     2  130535    7175  14506  4598
 3     3     3  247424   15663  31964  9589
 4     3     4  156871    7694  22884  7964
 5     3     5   64130    3241   6560  1752
 6     2     1  130706    7910  13792  4055
 7     2     2    7233     431   1001   380
 8     2     3   15868    1080   2695   701
 9     2     4   17707     888   3054   983
10     2     5    4039     209    487   114
11     1     1  163544    9862  19346  5552
12     1     2    9726     572   1430   439
13     1     3   20369    1382   3546  1011
14     1     4   21089    1052   3618  1281
15     1     5    4869     250    613   178
16     0     1  273944   17226  37730 11809
17     0     2   21504    1207   3421  1088
18     0     3   37666    2502   7565  2383
19     0     4   56730    2756  11345  3971
20     0     5    8601     461   1291   382
```

# Exploratory Data Analysis:




Linear Effect for Merit Rating, but quadratic effect for Class if considered as metric variables. Interaction effects present, especially when Merit=2 and 3.

# Likelihood analysis in Poisson regression

loglikelihood:

$$\ln l(\mathbf{y}, \boldsymbol{\beta}) = -\sum_{i=1}^{n} t_i \mu(\mathbf{x_i}, \boldsymbol{\beta}) + \sum_{i=1}^{n} y_i \ln(t_i \mu(\mathbf{x_i}, \boldsymbol{\beta})) + \text{ const. ind. of } \boldsymbol{\beta}$$

$$\Rightarrow s_j(\boldsymbol{\beta}) := \frac{\partial \ln l(\mathbf{y}, \beta)}{\partial \beta_j} = -\sum_{i=1}^{n} t_i e^{\mathbf{x_i}^t \beta} x_{ij} + \sum_{i=1}^{n} y_i \frac{e^{\mathbf{x_i}^t \beta}}{e^{\mathbf{x_i}^t \beta}} x_{ij}$$

$$= \sum_{i=1}^{n} (y_i - t_i e^{\mathbf{x_i}^t \beta}) x_{ij} = 0 \qquad j = 1, \ldots, p$$

$$\Rightarrow \mathbf{s}(\boldsymbol{\beta}) = \begin{pmatrix} s_1(\boldsymbol{\beta}) \\ \vdots \\ s_p(\boldsymbol{\beta}) \end{pmatrix} = X^t(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0} \qquad \boldsymbol{\mu} = \begin{pmatrix} t_i e^{\mathbf{x_i}^t \beta} \\ \vdots \\ t_i e^{\mathbf{x_i}^t \beta} \end{pmatrix}$$

score equations

MLE $\hat{\boldsymbol{\beta}}$ solves $\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$.

# Fisher information in Poisson regression

$$\frac{\partial^2 \ln l(\mathbf{y},\beta)}{\partial\beta_r\partial\beta_s} = \frac{\partial}{\partial\beta_r}\left(\sum_{i=1}^{n}(y_i - t_i e^{\mathbf{x_i}^t\beta})x_{is}\right)$$

$$= -\sum_{i=1}^{n} t_i e^{\mathbf{x_i}^t\beta} x_{is}x_{ir}$$

$$i_{rs} := E\left(-\frac{\partial^2 \ln l(\mathbf{y},\beta)}{\partial\beta_r\partial\beta_s}\right) = \sum_{i=1}^{n} t_i e^{\mathbf{x_i}^t\beta} x_{is}x_{ir}$$

$$\Rightarrow \quad I(\boldsymbol{\beta}) := (i_{rs})_{r,s=1,\ldots,p} = X^t D(\boldsymbol{\beta}) X \quad \text{where}$$

$$D(\boldsymbol{\beta}) := diag(\ldots, t_i e^{\mathbf{x_i}^t\beta}, \ldots)$$

$I(\hat{\boldsymbol{\beta}})^{-1} =$ estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$

# Poisson regression as GLM

$$
\begin{aligned}
P(Y_i = y_i) &= \exp\{-t_i\mu(\mathbf{x_i}, \boldsymbol{\beta}) + y_i \ln(t_i\mu(\mathbf{x_i}, \boldsymbol{\beta})) - \ln(y_i!)\} \\
&= exp\{y_i \underbrace{\ln(t_i\mu(\mathbf{x_i}, \boldsymbol{\beta}))}_{\theta_i} - \underbrace{t_i\mu(\mathbf{x_i}, \boldsymbol{\beta})}_{b(\theta_i)=e^{\theta_i}} - \underbrace{\ln(y_i!)}_{c(y_i,\phi)}\} \\
& a(\phi) = 1 \quad \phi = 1
\end{aligned}
$$

If $\mu(\mathbf{x_i}, \boldsymbol{\beta}) = e^{\mathbf{x_i}^t \beta}$

$$\Rightarrow \quad \mu_i = E(Y_i) = b'(\theta_i) = e^{\theta_i} = t_i\mu(\mathbf{x_i}, \boldsymbol{\beta}) = t_i e^{\mathbf{x_i}^t \beta} = e^{\mathbf{x_i}^t \beta + \ln(t_i)}$$

$$\eta_i = \boldsymbol{x_i^t \beta} = \ln(\mu_i) - \ln(t_i) = \theta_i - \ln(t_i) \qquad \ln(t_i) \text{ known offset}$$

$$\eta_i = g(\mu_i) \Rightarrow g(\mu_i) = \ln(\mu_i) - \ln(t_i) \qquad \text{extended link function}$$

If $t_i = 1 \ \forall i \Rightarrow \theta_i = \eta_i \Rightarrow g(\mu_i) = \ln(\mu_i)$ canonical link.

# Properties of the MLE

Since $\hat{\boldsymbol{\beta}}$ solves $\quad \mathbf{s}(\hat{\boldsymbol{\beta}}) = X^t(\mathbf{Y} - \hat{\boldsymbol{\mu}}) = \mathbf{0}$

$$\Rightarrow X^t\mathbf{Y} = X^t\hat{\boldsymbol{\mu}} \qquad \text{where} \quad \hat{\boldsymbol{\mu}} = (\ldots, t_i e^{\mathbf{x_i}^t\hat{\boldsymbol{\beta}}}, \ldots)^t$$

i) If $X$ contains intercept

$$\Rightarrow \sum_{i=1}^{n} Y_i = \underbrace{\mathbf{1_n}^t}_{\text{1st column of } X} \mathbf{Y} = \mathbf{1_n}^t\hat{\boldsymbol{\mu}} = \sum_{i=1}^{n} \hat{\mu}_i$$

ii) If

$$X = (\mathbf{D_1}, \mathbf{D_2}, \ldots, \mathbf{D_k}) \qquad D_{ij} = \begin{cases} 1 & i^{th} \text{ obs. has level } j \\ 0 & \text{otherwise} \end{cases}$$

$$j = 1, \ldots, k, \quad i = 1, \ldots, n, \quad k = p$$

$$\Rightarrow X^t\mathbf{Y} = (N_1, \ldots, N_k) \quad \text{where} \quad N_j = \text{ number of obs. with level } i$$

$\Rightarrow$ Fitted marginal totals $(X^t\hat{\boldsymbol{\mu}})$ are the same as the marginal totals $(X^t\mathbf{Y})$.

iii) If two factors $A$ with $I$ levels and $B$ with $J$ levels are considered with interaction model

$$Y \sim A * B$$

then the observed totals of each cell is the same as the fitted totals. If we only have a single observation for each cell, then we have a saturated model.

# Deviance analysis in Poisson regression

- After the EDA identifies important covariates one can use the partial deviance test to test for significance of individual or groups of covariates

- Goodnes of fit can be checked with the residual deviance test

- The deviance is given by

$$D = 2\sum_{i=1}^{n}\{y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\} \overset{a}{\sim} \chi^2_{n-p},$$

where $\hat{\mu}_i := t_i e^{\mathbf{x_i}^t \hat{\beta}}$.

**Remark:** The second term of $D$ will be zero if an intercept is used since $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\mu}_i$.

Pearson $\chi^2_P$ statistic for Poisson GLM's is given by

$$\chi^2_P = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \overset{a}{\sim} \chi^2_{n-p}$$

Note: " $\overset{a}{\sim}$ " assumes $n$ fixed, but $t_i \mu_i \to \infty \quad \forall i$.

# Poisson Regression:

## Class and Merit as factors

```
> f.main_Claims ~ offset(log(Insured)) + Merit + Class
> r.main_glm(f.main,family=poisson)
> summary(r.main,cor=F)
Call: glm(formula = Claims ~ offset(log(Insured)) +
        Merit + Class, family = poisson)
Deviance Residuals:
 Min 1Q Median  3Q Max
 -11 -3   -1.6 2.4  12

Coefficients:
            Value Std. Error t value
(Intercept) -2.04     0.0043    -472
     Merit1 -0.14     0.0072     -19
     Merit2 -0.22     0.0080     -28
     Merit3 -0.49     0.0045    -109
     Class2  0.30     0.0073      41
     Class3  0.47     0.0050      93
     Class4  0.53     0.0054      98
     Class5  0.22     0.0107      20

    Null Deviance: 33854 on 19 degrees of freedom

Residual Deviance: 580 on 12 degrees of freedom
```

Residual Deviance too large, not a good fit.

## With all interaction terms

```
> f.inter_Claims ~ offset(log(Insured))
                 + Merit * Class
> r.inter_glm(f.inter,family=poisson)
> summary(r.inter,cor=F)

Call: glm(formula = Claims ~ offset(log(Insured))
       + Merit * Class, family = poisson)
```

```
Coefficients:
                Value Std. Error  t value
 (Intercept) -1.9825     0.0051 -385.079
      Merit1 -0.1521     0.0088  -17.204
      Merit2 -0.2664     0.0100  -26.772
      Merit3 -0.5590     0.0056 -100.228
      Class2  0.1442     0.0179    8.074
      Class3  0.3772     0.0126   29.946
      Class4  0.3729     0.0107   34.830
      Class5  0.0860     0.0283    3.039
Merit1Class2  0.0733     0.0327    2.241
Merit2Class2  0.1270     0.0373    3.407
Merit3Class2  0.2003     0.0198   10.110
Merit1Class3  0.0092     0.0222    0.413
Merit2Class3  0.0987     0.0245    4.022
Merit3Class3  0.1178     0.0139    8.442
Merit1Class4 -0.0012     0.0210   -0.056
Merit2Class4  0.1184     0.0227    5.220
Merit3Class4  0.2436     0.0128   19.080
Merit1Class5 -0.0237     0.0498   -0.475
Merit2Class5  0.0474     0.0541    0.876
Merit3Class5  0.1756     0.0310    5.672

    Null Deviance: 33854 on 19 degrees of freedom

Residual Deviance:  0 on 0 degrees of freedom
```

This model is the saturated model since only one observation per cell. Interaction effects when Merit=3 or 2 are very significant.

## With some interaction terms

```
> f.inter1_Claims ~ offset(log(Insured)) + Merit
+ Class + Class:(Merit == 3) + Class:(Merit == 2)
> r.inter1_glm(f.inter1,family=poisson)
> summary(r.inter1,cor=F)

Call: glm(formula = Claims ~ offset(log(Insured)) +
Merit + Class + Class:(Merit == 3) +
Class:(Merit == 2), family = poisson)
```

```
Coefficients: (4 not defined because of singularities)
                   Value Std. Error t value
     (Intercept) -1.9839     0.0048 -409.37
          Merit1 -0.1479     0.0072  -20.58
          Merit2 -0.2100     0.0508   -4.13
          Merit3 -0.3744     0.0261  -14.32
          Class2  0.1655     0.0150   11.06
          Class3  0.3802     0.0104   36.67
          Class4  0.3731     0.0092   40.50
          Class5  0.0784     0.0233    3.36
      Merit == 3      NA         NA      NA
      Merit == 2      NA         NA      NA
Merit == 3Class1 -0.1832     0.0265   -6.93
Merit == 3Class2 -0.0043     0.0309   -0.14
Merit == 3Class3 -0.0685     0.0283   -2.42
Merit == 3Class4  0.0602     0.0281    2.15
Merit == 3Class5      NA         NA      NA
Merit == 2Class1 -0.0550     0.0517   -1.06
Merit == 2Class2  0.0507     0.0615    0.82
Merit == 2Class3  0.0407     0.0551    0.74
Merit == 2Class4  0.0633     0.0545    1.16
Merit == 2Class5      NA         NA      NA

    Null Deviance: 33854 on 19 degrees of freedom

Residual Deviance: 5.5 on 4 degrees of freedom
```

The Residual Deviance is now only 5.5 on 4 df with a p-value of .24, reasonable good fit.

## Class and Merit as metric variables

```
> Merit.m_as.numeric(Merit)
> Class.m_as.numeric(Class)
> f.main.m_Claims ~ offset(log(Insured))+Merit.m + poly(Class.m,2)
> r.main.m_glm(f.main.m,family=poisson)
> summary(r.main.m,cor=F)

Call: glm(formula = Claims ~ offset(log(Insured))
       + Merit.m + poly(Class.m, 2), family = poisson)

Coefficients:
                  Value Std. Error t value
      (Intercept) -1.53     0.0051    -302
          Merit.m -0.17     0.0014    -121
poly(Class.m, 2)1  0.48     0.0134      36
poly(Class.m, 2)2 -0.64     0.0118     -54

    Null Deviance: 33854 on 19 degrees of freedom

Residual Deviance: 1000 on 16 degrees of freedom
```

## With interaction

```
> f.inter.m_Claims ~ offset(log(Insured))
                   + Merit.m*poly(Class.m,2)
> r.inter.m_glm(f.inter.m,family=poisson)
> summary(r.inter.m)
Call: glm(formula = Claims ~ offset(log(Insured))
       + Merit.m * poly(Class.m, 2), family =
        poisson)
Deviance Residuals:
  Min    1Q Median   3Q Max
 -8.3 -2.9  -0.99 2.8  12
```

```
Coefficients:
                                Value Std. Error
               (Intercept) -1.601     0.0066
                   Merit.m -0.145     0.0020
        poly(Class.m, 2)1  0.161     0.0390
        poly(Class.m, 2)2 -0.476     0.0362
Merit.mpoly(Class.m, 2)1  0.102     0.0112
Merit.mpoly(Class.m, 2)2 -0.048     0.0103
                          t value
               (Intercept)  -241.4
                   Merit.m   -73.9
        poly(Class.m, 2)1     4.1
        poly(Class.m, 2)2   -13.2
Merit.mpoly(Class.m, 2)1     9.0
Merit.mpoly(Class.m, 2)2    -4.6

    Null Deviance: 33854 on 19 degrees of freedom
Residual Deviance: 580 on 14 degrees of freedom
```

Residual deviance is still too high the models with Merit and Class as factors fit better. Note this is not a saturated model.

# Residuals in Poisson regression

**Pearson residuals:** $r_i^P := \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$

**Deviance residuals:** $r_i^D := sign(y_i.\hat{\mu}_i)[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)]^{1/2}$

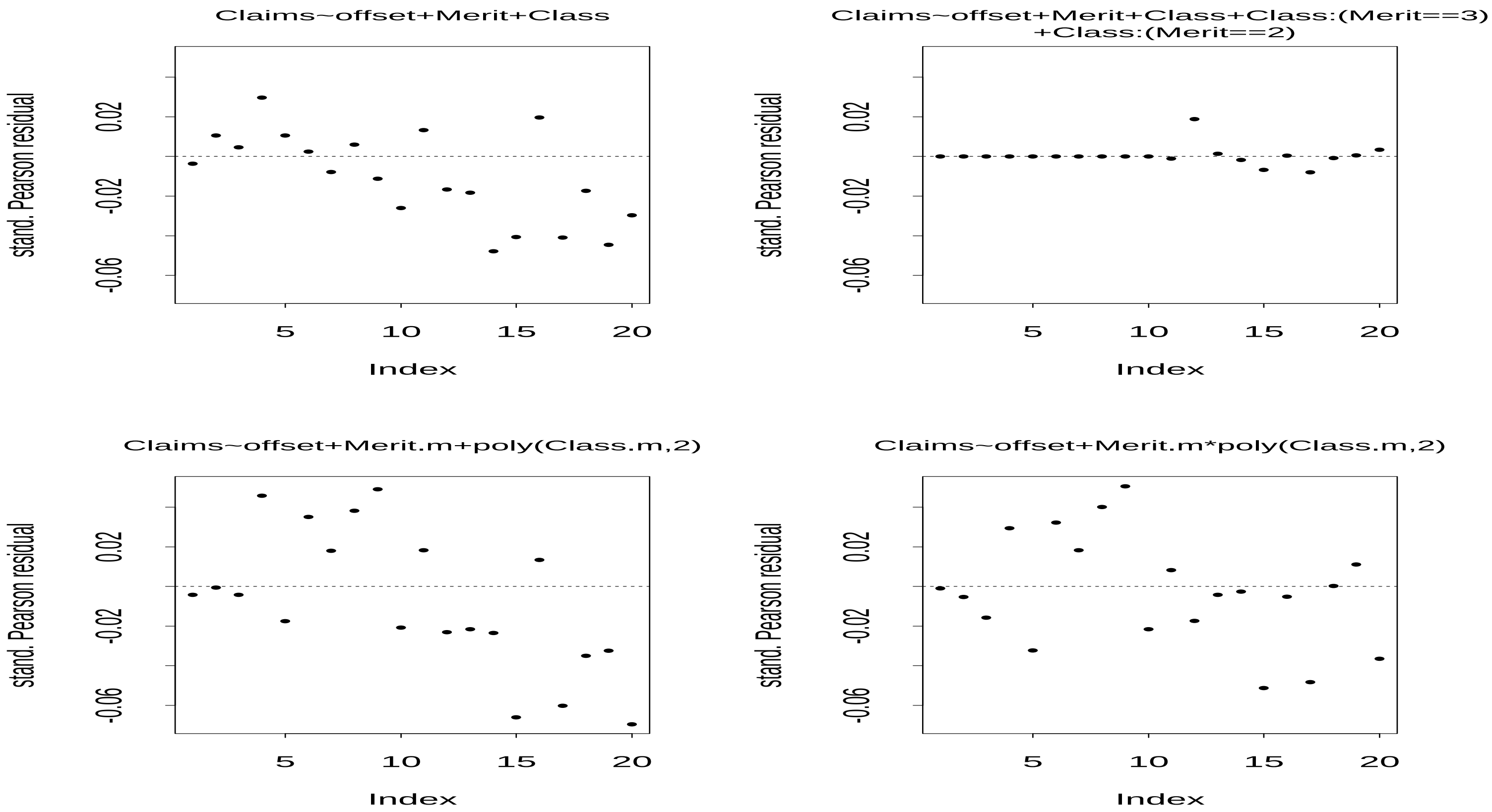
resid(r.object,type="deviance")
resid(r.object,type="pearson")  in Splus

If $t_i \neq const \quad \forall\, i$, it is better to consider the standardized response $Y_i^* = Y_i/t_i$

$$\Rightarrow r_i^{P*} := \frac{y_i/t_i - \hat{\mu}_i/t_i}{\sqrt{\hat{\mu}_i/t_i}} = \frac{1}{t_i}\sqrt{t_i}r_i^P = \frac{1}{\sqrt{t_i}}r_i^P$$ standardized Pearson residuals

**Interpretation:**
$\hat{\mu}_i = t_i e^{\mathbf{x_i}^t \hat{\beta}}$ is the estimated number of events in 0 to $t_i$ units $\Rightarrow \frac{\hat{\mu}_i}{t_i} =$ the estimated number of events in 0 to 1 units.

# Residual Analysis




Residuals for model

```
Claims~offset(log(Insured))+Merit+Class+Class:(Merit==3)+Class:(Merit==2)
```

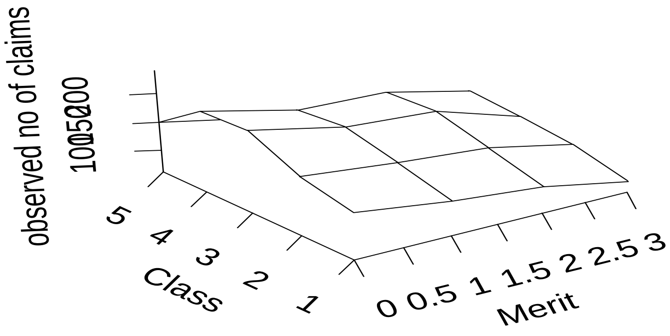
look best.

# Interpretation

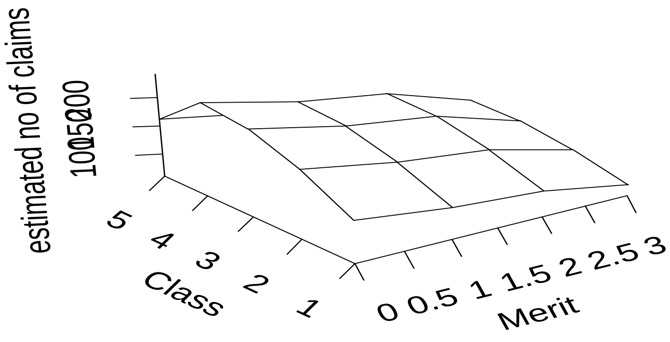
The estimated number of claims per 1000 earned car years for the main effects only and the interaction model are given by:

```
> fit.main_tapply((fitted(r.main)/Insured)*1000,list(Merit,Class),mean)
> fit.main
          1        2        3        4         5
0 130.58435 176.2405 208.7369 220.9363 161.99570
1 113.77928 153.5599 181.8742 192.5037 141.14826
2 104.72520 141.3402 167.4014 177.1851 129.91627
3  79.76373 107.6515 127.5010 134.9526  98.95045

> fit.inter1_tapply((fitted(r.inter1)/Insured)*1000,list(Merit,Class),mean)
> fit.inter1
          1        2        3        4        5
0 137.53150 162.2798 201.1603 199.7208 148.7432
1 118.62295 139.9687 173.5037 172.2622 128.2932
2 105.51926 138.3935 169.8387 172.4742 120.5744
3  78.74866 111.1273 129.1871 145.8778 102.2922

> fit.obs <- tapply((Claims/Insured) * 1000,list(Merit, Class), mean)
> fit.obs
          1        2        3        4        5
0 137.72888 159.0867 200.8443 199.9824 150.0988
1 118.29233 147.0286 174.0881 171.5586 125.8985
2 105.51926 138.3935 169.8387 172.4742 120.5744
3  78.74866 111.1273 129.1871 145.8778 102.2922
```

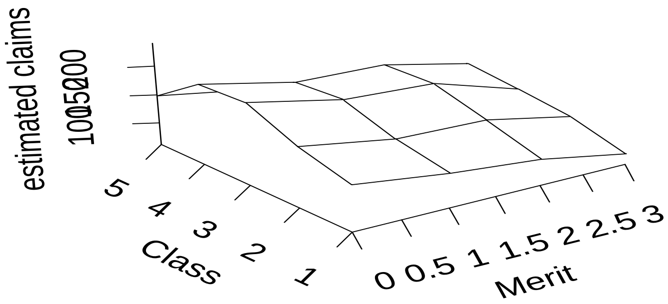
Observed Claims




Expected Claims (Main Effect




Expected Claims (With Interact



# References


Bailey, R. and L. Simon (1960). Two studies in automobile insurance ratemaking. *ASTIN Bulletin V.1, N.4*, 192–217.

Cameron, A. and P. Trivedi (1998). *Regression analysis of count data*. Cambridge University Press.

Winkelmann, R. (1997). *Econometric analysis of count data* (Second, revised and enlarged ed.). Berlin: Springer-Verlag.