Modelling Severity of Services

Swiss Paraplegic Forschrung

2023-04-06

Modelling Framework

The next model aims to explain the severity of services in terms of the total number hours of services that a patient from the clinic receives and their respective determinants. On one hand the length of stay defines the time that an individual facing a lesion in the spinal cord injury spends in the clinic independently of the service received. On the other hand we consider the intensity of care related to each service. The available services that we consider in the analysis are the following:

- Physioteraphy
- Ergotherapy
- Logopaedics
- Intensive Care Nursing
- Non medical therapies
- Psychology
- Anaesthesia
- Nursing
- Imaging
- Laboratory

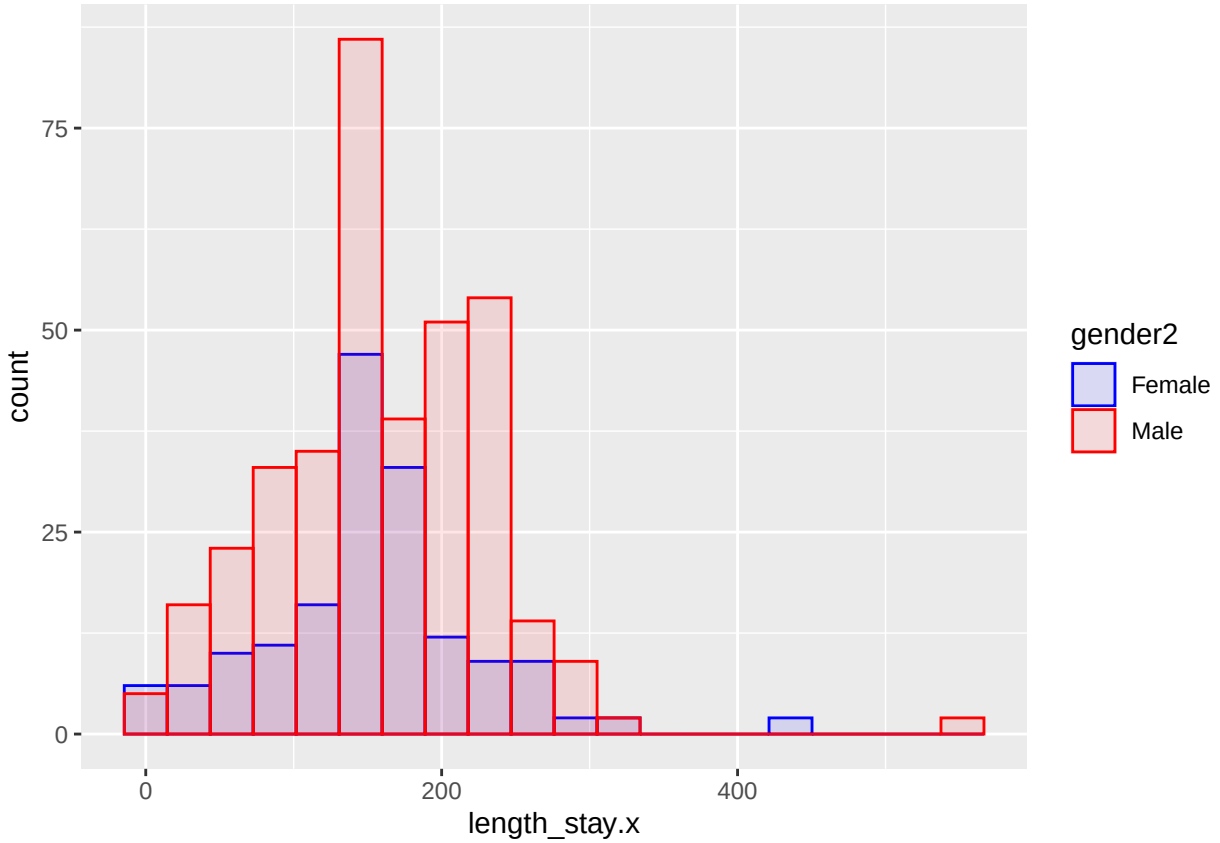
The modelling was based in the paper Modelling the burden of longterm care for institutionalised elderly based on care duration and intensity written by Martin Bladt and colleagues (Bladt et al. 2023). The idea is to consider that the *Intensity of Service* (total number of hours of service received) can be modeled by multiplying the duration or *lenght of stay* that is the time spent in the clinic, with the *intensity of care* that is the amount of service received expressed in minutes or swiss francs. This approach is well established in actuarial practice for example we can find similar ideas in (Frees, Gao, and Rosenberg 2011). These ideas are based in the well known Compound Poisson Risk model (Panjer and Willmot 1983). We will decompose the analysis in two stages, the first is the modeling of Lenght of stay and then the modeling of the intensity of care. This two stage approach allows us to investigate the time spent in the institution and the amount of care received.

Length of Stay

The length of stay was constructed by the difference between two date type variables, more specifically we considered the difference between the date of entry to the clinic and the leaving date. The next histogram shows main features of this variable:

```
pacman::p_load(knitr,reshape,dplyr,ggstatsplot,
               ROSE,ggsci,grDevices,VGAM,jtools,
               RColorBrewer,ggpubr,fishmethods,
               ggplot2,ggsci,margins,car,dplyr,scales,reshape2,varhandle,did)
colores<-get_palette(palette = "lancet",29)

dataset<-readRDS("tab_varnew.rds")
dataset$gender2<-as.factor(dataset$gender2)
ggplot(dataset,aes(x=length_stay.x,color=gender2,fill=gender2))+
  geom_histogram(alpha=0.1,bins = 20,position="identity")+
  scale_color_manual(values=c("blue","red"))+ scale_fill_manual(values=c("blue","red"))
```



- In the histogram presented above, we can see that there is a difference in the density distributions of the length of stay by gender. However, the shape is quite similar. Moreover for both cases we observe extreme values that possible are due to cases with complications that requires more medical services.
- We observe that in the dataset there were more male patients than female patients with a higher value of length of stay.
- Based in the shape of the histogram we could consider that the normal distribution is a good candidate for fitting the observed durations.

```
dataset|>group_by(PID)|>
  dplyr::summarize(N0=n(),gender=first(gender2),los=mean(length_stay.x))|>
  group_by(gender)|>dplyr::summarize(N=n(),mlos=mean(los))|>
  as.data.frame()|>kable()
```

gender	N	mlos
Female	122	145.2254
Male	276	153.5616

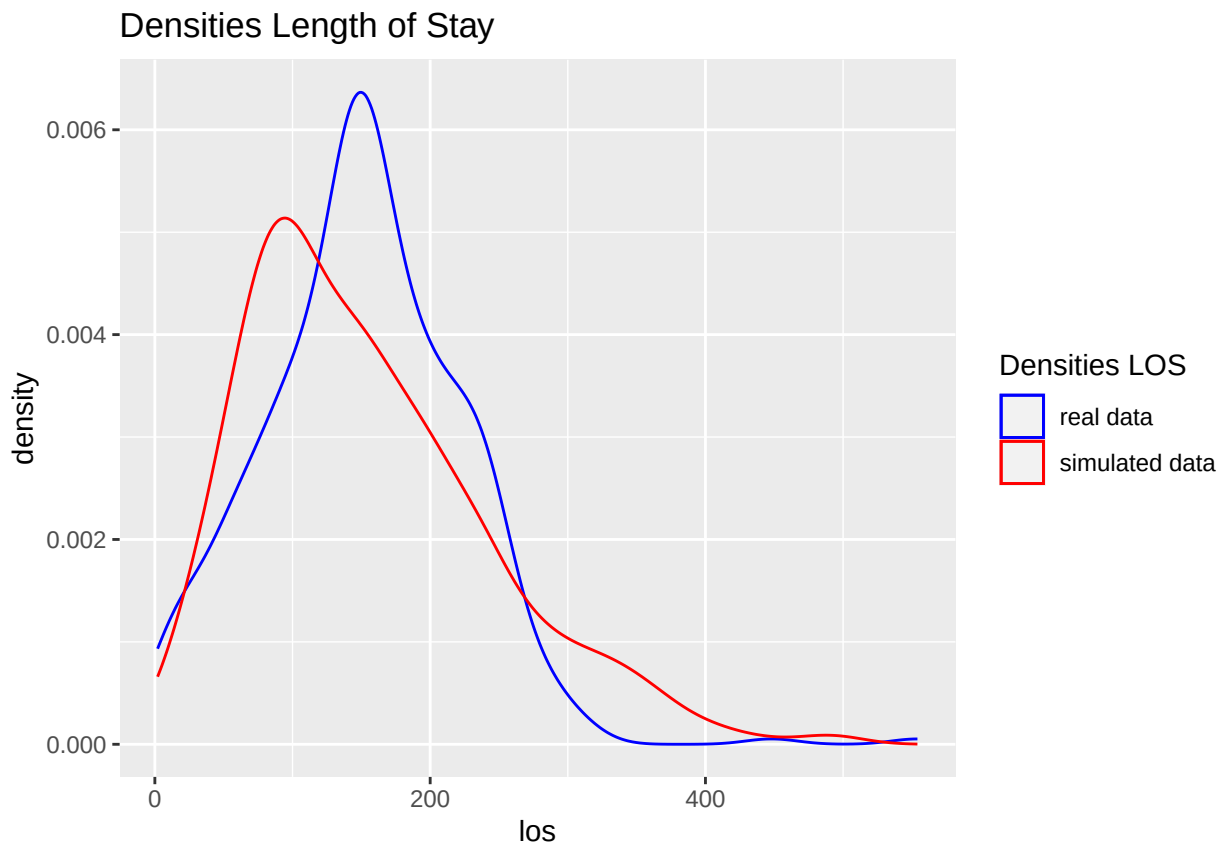
Length of stay model.

To model the length of stay that we are going to call D, we first asses its distribution by fitting the observed durations with distributions commonly established in the literature. We consider a zero truncated negative binomial distribution that is a negative binomial distribution conditional on it taking positive values. That characteristic is well suited for our data due to there are not cases with null length of stay values in the sample. Details about this approach could we found in some literature as Yang and Berdine (2015). The computations were performed with the library VGAM for vector generalized and additive models technical details about this package could be found in (Yee 2015).

```
los<-dataset|>group_by(id)|>summarize(los=mean(length_stay.x))
fitdistrplus::fitdist(los$los, "nbinom")
```

```
## Fitting of the distribution ' nbinom ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## size   3.15559  0.2225472
## mu   150.63285  4.2842447
```

```
los<-los|>mutate(simlos=rnbinom(400,size = 3.15559,mu=150.63))
ggplot(los,aes(x=los))+geom_density(aes(color="real data"))+
  geom_density(aes(x=simlos,color="simulated data"))+
  labs(title = "Densities Length of Stay")+
  scale_color_manual(name='Densities LOS',
                     breaks=c('real data', 'simulated data'),
                     values=c('real data'='blue', 'simulated data'='red'))
```



Regression Estimation

```
## `summarise()` has grouped output by 'id'. You can override using the `.groups`
## argument.
```

The log likelihood of the model is -1858.561. From the summary of the zero-truncated negative binomial regression model, the variable sex2 corresponding to female patients, has an estimated coefficient 0.043372 that is not significant, because the p-value is $0.4569 > 0.05$ which means that the log count of days in the clinic is unperceivable higher compared to male patients by 0.04 when the other covariates are the same.

Regarding the age group we have that for age groups $[45, 60]$ and > 75 the estimated coefficients are 0.16 and -0.20 respectively, indicating that the log count of days in the clinic are higher compared to the age group $[10, 31]$ and lower for the age group > 75 by -0.20 when the other variables are the same.

Additionally, we have that the estimated coefficient of the group Paraplegia, ASIA A,B,C,D is -0.2177 suggesting that log counts of lenght of stay for patients in this group is less than the group $C1 - C4, ASIA A, B, C, D$ by 0.21 when the other covariates keep constant. Regarding the variable stay after leaving, indicates that for patients with discharge destination Home or Nursing home the log of the length of stay is significantly higher by 0.89 and 0.74 respectively. Finally, we observe the SCIM value has a negative significant coefficient indicating that for each unit the log of LOS decreases -0.015616 for each SCIM increase when the other variables keep the same.

Moreover, we observe two estimated coefficients for two intercepts, the first one ((Intercept):1) is 4.71, that can be interpreted as the typical intercept indicating the log count of the stay when all the predictos are equal to zero. The value of the second intercept ((intercept):2) is the overdispersion parameter (alpha) that is equals to 1.5357.

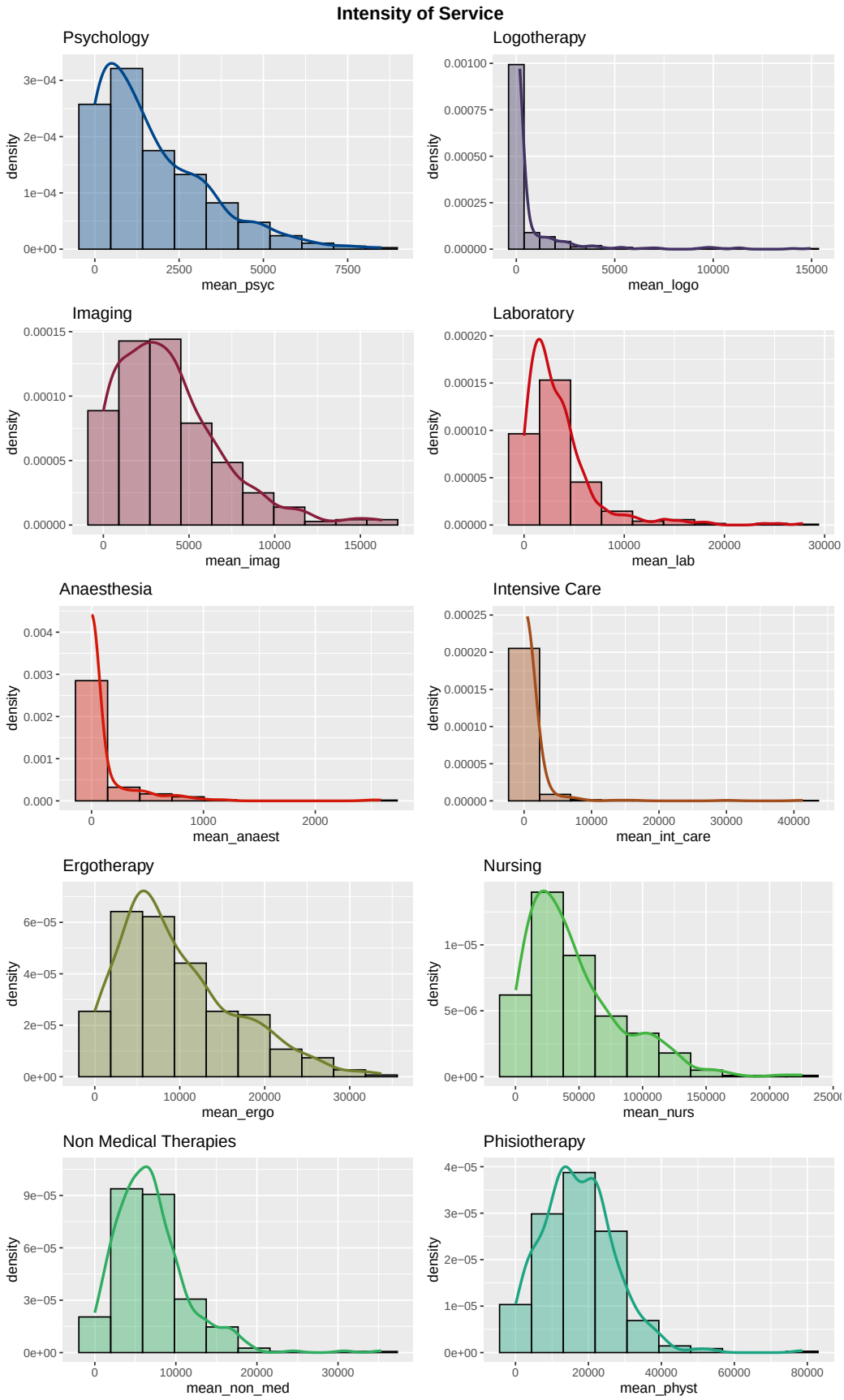
```
knitr::kable(summ1@coef3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept):1	4.7110795	0.2735724	17.2205950	0.0000000
(Intercept):2	1.5357281	0.0777162	19.7607209	0.0000000
sex2	0.0433721	0.0583029	0.7439100	0.4569309
age_group]31,45]	0.1011258	0.0956686	1.0570431	0.2904919
age_group]45,60]	0.1600932	0.0863500	1.8540033	0.0637387
age_group]60,75]	0.0072668	0.0891277	0.0815325	0.9350185
age_group>75	-0.2016655	0.1156444	-1.7438418	0.0811867
group1C5-C8, ASIA A,B,C,D	-0.0487874	0.0864570	-0.5642969	0.5725521
group1Paraplegia, ASIA A,B,C,D	-0.2177730	0.0759398	-2.8677074	0.0041346
etiologySickness	0.0332351	0.0675154	0.4922599	0.6225356
main_cost2	0.2189106	0.2848586	0.7684888	0.4421969
main_cost3	-0.8815064	0.3607729	-2.4433831	0.0145503
main_cost4	0.1443412	0.0813175	1.7750334	0.0758924
main_cost5	0.1500365	0.2909609	0.5156586	0.6060929
stay_aft_leav1	0.8919942	0.2477738	3.6000341	0.0003182
stay_aft_leav4	0.4783875	0.3732699	1.2816129	0.1999785
stay_aft_leav5	-0.2954388	0.3224019	-0.9163682	0.3594738
stay_aft_leav6	0.2101691	0.2881115	0.7294714	0.4657133
stay_aft_leavNurse	0.7493095	0.2550401	2.9380071	0.0033033
ins_classPrivate	-0.0449046	0.0855366	-0.5249752	0.5996004
ins_classSemi-Private	-0.1055241	0.0851176	-1.2397452	0.2150697
SCIM	-0.0156156	0.0013624	-11.4620583	0.0000000

Intensity of Care

We define the intensity of care as the amount of time spent in an specific service from the services mentioned above. First we show the histograms of the intensity of care for each service in the sample.

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```


Intensity of Service
Psychology
density
mean_psyc
Logotherapy
density
mean_logo
Imaging
density
mean_imag
Laboratory
density
mean_lab
Anaesthesia
density
mean_anaest
Intensive Care
density
mean_int_care
Ergotherapy
density
mean_ergo
Nursing
density
mean_nurs
Non Medical Therapies
density
mean_non_med
Phisiotherapy
density
mean_physt

Services	Units
Anaesthesia	min
Intensive Care	min
Imaging	CHF
Laboratory	CHF
Physiotherapy	min
Ergotherapy	min
Logotheraphy	min
Non Medical Therapies	min
Nursing	min
Psychology	min

Regression Estimation Severity of Service

Now we perform estimations of the intensity of service for each service described above, More specifically given a set of $n = 10$ services expressed as follows:

We are going to proceed by considering some of the most common models in the literature for severity analysis. First since we note we are dealing with positive values, greater or equal to zero..

```
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 4.2.3
```

```
library(survival)

m1_sev<-glm(mean_logo~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_logo>0),family= gaussian(link='log'))


m2_sev<-glm(mean_psyc~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_psyc>0),family= gaussian(link='log'))

m3_sev<-glm(mean_imag~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_imag>0),family= gaussian(link='log'))

m4_sev<-glm(mean_lab~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_lab>0),family= gaussian(link='log'))


m5_sev<-glm(mean_nurs~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_nurs>0),family= gaussian(link='log'))

m6_sev<-glm(mean_anaest~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_anaest>0),family= gaussian(link='log'))

m7_sev<-glm(mean_int_care~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_int_care>0),family= gaussian(link='log'))
```

```
## Warning: glm.fit: algorithm did not converge
```

```r
m8_sev<-glm(mean_ergo~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_ergo>0),family= gaussian(link='log'))

m9_sev<-glm(mean_non_med~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_non_med>0),family= gaussian(link='log'))


m10_sev<-glm(mean_physt~sex+age_group+group1+etiology+main_cost+stay_aft_leav+ins_class+SCIM,
         data=data_regression|>filter(mean_physt>0),family= gaussian(link='log'))


colnames(data_regression)
```

```
##  [1] "id"            "group1"        "mean_logo"     "mean_psyc"
##  [5] "mean_imag"     "mean_lab"      "mean_nurs"     "mean_anaest"
##  [9] "mean_int_care" "mean_ergo"     "mean_physt"    "mean_non_med"
## [13] "mean_los"      "N"             "sex"           "etiology"
## [17] "age_group"     "injury_group"  "stay_aft_leav" "main_cost"
## [21] "group_level"   "ins_class"     "ins_type"      "SCIM"
## [25] "age"
```

```r
summary(m1_sev)
```

```
##
## Call:
## glm(formula = mean_logo ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_logo > 0))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3515.0   -116.3    134.4    685.9   7043.6
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.74537    1.52385   3.114  0.00241 **
## sex2                             0.06630    0.46713   0.142  0.88742
## age_group]31,45]                 1.73065    0.96687   1.790  0.07652 .
## age_group]45,60]                 2.07549    0.86006   2.413  0.01765 *
## age_group]60,75]                 2.55487    0.95137   2.685  0.00849 **
## age_group>75                     2.64085    1.05965   2.492  0.01436 *
## group1C5-C8, ASIA A,B,C,D       -0.62741    0.37830  -1.658  0.10038
## group1Paraplegia, ASIA A,B,C,D -1.14869    0.53854  -2.133  0.03540 *
## etiologySickness                 0.60647    0.23957   2.531  0.01293 *
## main_cost2                      10.41072    1.83907   5.661 1.47e-07 ***
## main_cost3                       0.30853    0.71314   0.433  0.66622
## main_cost4                       1.24780    0.47383   2.633  0.00981 **
## stay_aft_leav1                   1.05247    1.23524   0.852  0.39625
## stay_aft_leav4                   5.40274    4.00783   1.348  0.18072
## stay_aft_leav5                   2.60981    1.27725   2.043  0.04368 *
## stay_aft_leav6                  -1.36972    5.26615  -0.260  0.79533
## stay_aft_leavNurse               1.40798    1.16992   1.203  0.23166
```

```
## ins_classPrivate               -1.58306    0.62506  -2.533  0.01289 *
## ins_classSemi-Private           0.19771    0.58638   0.337  0.73670
## SCIM                           -0.11206    0.02246  -4.988 2.61e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2672105)
##
##     Null deviance: 559677615  on 118  degrees of freedom
## Residual deviance: 264434399  on  99  degrees of freedom
##   (25 observations deleted due to missingness)
## AIC: 2118.8
##
## Number of Fisher Scoring iterations: 16
```

```
summary(m2_sev)
```

```
##
## Call:
## glm(formula = mean_psyc ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_psyc > 0))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2913.9  -1016.6   -311.2    753.4   5730.2
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      7.016255   1.381766   5.078 7.07e-07 ***
## sex2                             0.136764   0.099978   1.368 0.172454
## age_group]31,45]                -0.317677   0.157263  -2.020 0.044353 *
## age_group]45,60]                -0.207720   0.134530  -1.544 0.123732
## age_group]60,75]                -0.515905   0.140793  -3.664 0.000298 ***
## age_group>75                    -0.703511   0.202007  -3.483 0.000578 ***
## group1C5-C8, ASIA A,B,C,D       -0.011940   0.135270  -0.088 0.929730
## group1Paraplegia, ASIA A,B,C,D -0.026908   0.122997  -0.219 0.826991
## etiologySickness                 0.031330   0.113855   0.275 0.783386
## main_cost2                      -0.115669   0.434303  -0.266 0.790183
## main_cost3                      -0.994911   2.078586  -0.479 0.632571
## main_cost4                      -0.176701   0.134631  -1.312 0.190456
## main_cost5                       0.117876   0.478817   0.246 0.805725
## stay_aft_leav1                   1.272717   1.370185   0.929 0.353777
## stay_aft_leav4                   1.034202   1.454229   0.711 0.477585
## stay_aft_leav5                   0.560278   1.486529   0.377 0.706537
## stay_aft_leav6                   1.017488   1.418347   0.717 0.473753
## stay_aft_leavNurse               1.341805   1.371931   0.978 0.328916
## ins_classPrivate                -0.179229   0.187605  -0.955 0.340241
## ins_classSemi-Private            0.008541   0.148076   0.058 0.954047
## SCIM                            -0.011707   0.002939  -3.984 8.71e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2246945)
```

```
##
##     Null deviance: 717842341  on 294  degrees of freedom
## Residual deviance: 615653706  on 274  degrees of freedom
##   (56 observations deleted due to missingness)
## AIC: 5173.8
##
## Number of Fisher Scoring iterations: 9
```

```
summary(m3_sev)
```

```
##
## Call:
## glm(formula = mean_imag ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_imag > 0))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5730.9  -1770.3   -470.8   1204.8  11734.9
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.829975   0.396500  22.270  < 2e-16 ***
## sex2                           -0.097405   0.100287  -0.971  0.33222
## age_group]31,45]                0.130121   0.173530   0.750  0.45395
## age_group]45,60]                0.397776   0.143566   2.771  0.00596 **
## age_group]60,75]                0.055777   0.156030   0.357  0.72099
## age_group>75                    0.048024   0.188963   0.254  0.79956
## group1C5-C8, ASIA A,B,C,D      -0.169447   0.115172  -1.471  0.14231
## group1Paraplegia, ASIA A,B,C,D -0.090545   0.103589  -0.874  0.38280
## etiologySickness                0.017457   0.099477   0.175  0.86082
## main_cost2                     -0.246342   0.825779  -0.298  0.76568
## main_cost3                      0.410228   0.424236   0.967  0.33436
## main_cost4                     -0.048039   0.119858  -0.401  0.68886
## main_cost5                      0.398555   0.344567   1.157  0.24835
## stay_aft_leav1                 -0.074898   0.349853  -0.214  0.83063
## stay_aft_leav4                 -1.117317   1.267194  -0.882  0.37866
## stay_aft_leav5                 -0.778030   0.534096  -1.457  0.14627
## stay_aft_leav6                 -0.085050   0.420104  -0.202  0.83971
## stay_aft_leavNurse             -0.119015   0.355792  -0.335  0.73824
## ins_classPrivate                0.136886   0.124980   1.095  0.27431
## ins_classSemi-Private          -0.102427   0.151541  -0.676  0.49964
## SCIM                           -0.015099   0.002834  -5.328    2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 8233245)
##
##     Null deviance: 2979039101  on 310  degrees of freedom
## Residual deviance: 2387547430  on 290  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 5857.1
##
## Number of Fisher Scoring iterations: 8
```

```
summary(m4_sev)
```

```
##
## Call:
## glm(formula = mean_lab ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_lab > 0))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -8007.9  -1412.6   -339.1    847.4  17094.8
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      9.845568   0.284551  34.600  < 2e-16 ***
## sex2                            -0.125091   0.107710  -1.161  0.24639
## age_group]31,45]                -0.034444   0.243003  -0.142  0.88737
## age_group]45,60]                 0.342946   0.188116   1.823  0.06926 .
## age_group]60,75]                 0.131575   0.196721   0.669  0.50410
## age_group>75                    -0.037224   0.237305  -0.157  0.87546
## group1C5-C8, ASIA A,B,C,D       -0.141907   0.129734  -1.094  0.27488
## group1Paraplegia, ASIA A,B,C,D   0.010948   0.124771   0.088  0.93014
## etiologySickness                 0.242509   0.119015   2.038  0.04244 *
## main_cost2                       0.440997   1.744719   0.253  0.80062
## main_cost3                       0.428252   0.392743   1.090  0.27638
## main_cost4                       0.005708   0.158104   0.036  0.97123
## main_cost5                       0.329399   0.586644   0.561  0.57487
## stay_aft_leav1                  -1.078522   0.160867  -6.704 9.61e-11 ***
## stay_aft_leav4                  -1.352623   0.907873  -1.490  0.13728
## stay_aft_leav5                  -1.894043   0.578135  -3.276  0.00117 **
## stay_aft_leav6                  -0.939438   0.237541  -3.955 9.50e-05 ***
## stay_aft_leavNurse              -1.017577   0.169684  -5.997 5.61e-09 ***
## ins_classPrivate                -0.031497   0.182798  -0.172  0.86331
## ins_classSemi-Private           -0.168486   0.169809  -0.992  0.32187
## SCIM                            -0.028081   0.003986  -7.045 1.21e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 9678453)
##
##     Null deviance: 4419941347  on 329  degrees of freedom
## Residual deviance: 2990525724  on 309  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 6267
##
## Number of Fisher Scoring iterations: 10
```

```
summary(m5_sev)
```

```
##
## Call:
## glm(formula = mean_nurs ~ sex + age_group + group1 + etiology +
```

```
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_nurs > 0))
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -49823  -11806   -2605    9876   65535
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    10.420639   0.608203  17.133  < 2e-16 ***
## sex2                           -0.075302   0.061857  -1.217   0.2244
## age_group]31,45]                0.003778   0.106928   0.035   0.9718
## age_group]45,60]                0.213525   0.087060   2.453   0.0147 *
## age_group]60,75]                0.075298   0.091613   0.822   0.4117
## age_group>75                   -0.047543   0.109786  -0.433   0.6653
## group1C5-C8, ASIA A,B,C,D      -0.017135   0.057943  -0.296   0.7676
## group1Paraplegia, ASIA A,B,C,D -0.276795   0.064793  -4.272 2.57e-05 ***
## etiologySickness               -0.052036   0.056653  -0.919   0.3591
## main_cost2                     -0.188008   0.773190  -0.243   0.8080
## main_cost3                     -0.213602   0.306754  -0.696   0.4867
## main_cost4                      0.011587   0.072865   0.159   0.8738
## main_cost5                      0.455831   0.281908   1.617   0.1069
## stay_aft_leav1                  1.227710   0.599238   2.049   0.0413 *
## stay_aft_leav4                  1.095207   0.719324   1.523   0.1289
## stay_aft_leav5                  0.597389   0.631820   0.946   0.3451
## stay_aft_leav6                  0.924880   0.616116   1.501   0.1343
## stay_aft_leavNurse              1.342486   0.599410   2.240   0.0258 *
## ins_classPrivate               -0.010739   0.087059  -0.123   0.9019
## ins_classSemi-Private          -0.106902   0.090160  -1.186   0.2366
## SCIM                           -0.028105   0.002218 -12.674  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 473160640)
##
##     Null deviance: 4.0925e+11  on 336  degrees of freedom
## Residual deviance: 1.4952e+11  on 316  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 7710.2
##
## Number of Fisher Scoring iterations: 8
```

```
summary(m6_sev)
```

```
##
## Call:
## glm(formula = mean_anaest ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_anaest > 0))
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -620.1  -160.5    10.2   239.7  1518.3
##
```

```
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     7.00011    0.69834  10.024 2.97e-13 ***
## sex2                           -0.03670    0.36790  -0.100  0.92096
## age_group]31,45]               -0.80819    0.75448  -1.071  0.28955
## age_group]45,60]               -0.63454    0.63085  -1.006  0.31964
## age_group]60,75]               -1.63040    0.70427  -2.315  0.02503 *
## age_group>75                   -2.72010    1.28748  -2.113  0.03996 *
## group1C5-C8, ASIA A,B,C,D      -0.55897    0.40670  -1.374  0.17583
## group1Paraplegia, ASIA A,B,C,D  0.43232    0.32742   1.320  0.19310
## etiologySickness                0.62051    0.42800   1.450  0.15375
## main_cost4                     -0.60741    0.57423  -1.058  0.29556
## main_cost5                     -0.66231    2.53989  -0.261  0.79542
## stay_aft_leav4                 -1.90053   17.90676  -0.106  0.91593
## stay_aft_leav6                  0.57013    0.59507   0.958  0.34292
## stay_aft_leavNurse              0.77378    0.31150   2.484  0.01661 *
## ins_classPrivate                1.28894    0.41392   3.114  0.00314 **
## ins_classSemi-Private          -0.06843    0.99582  -0.069  0.94551
## SCIM                           -0.02311    0.01452  -1.591  0.11831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 200161.2)
##
##     Null deviance: 13867310  on 63  degrees of freedom
## Residual deviance:  9406420  on 47  degrees of freedom
##   (17 observations deleted due to missingness)
## AIC: 979.1
##
## Number of Fisher Scoring iterations: 23
```

```
summary(m7_sev)
```

```
##
## Call:
## glm(formula = mean_int_care ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_int_care > 0))
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -189.8    15.0   152.0   797.0  7971.0
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     108.996    181.689   0.600    0.550
## sex2                            -67.400    128.307  -0.525    0.600
## age_group]31,45]                -63.168    107.469  -0.588    0.558
## age_group]45,60]                -18.760     37.504  -0.500    0.618
## age_group]60,75]                -46.933     80.225  -0.585    0.560
## age_group>75                    -20.382     37.224  -0.548    0.585
## group1C5-C8, ASIA A,B,C,D         7.837     16.123   0.486    0.628
## group1Paraplegia, ASIA A,B,C,D   17.159     31.835   0.539    0.591
## etiologySickness                  8.129     10.917   0.745    0.458
```

```
## main_cost2                      106.091    187.853   0.565    0.573
## main_cost3                       -8.408     16.617  -0.506    0.614
## main_cost4                       27.879     48.591   0.574    0.567
## main_cost5                       81.034    150.264   0.539    0.591
## stay_aft_leav1                  -76.408    139.253  -0.549    0.584
## stay_aft_leav4                  -30.174     70.960  -0.425    0.671
## stay_aft_leav5                 -125.227    224.781  -0.557    0.578
## stay_aft_leav6                  -82.821    144.591  -0.573    0.568
## stay_aft_leavNurse              -63.413    112.255  -0.565    0.573
## ins_classPrivate                 11.037     10.851   1.017    0.311
## ins_classSemi-Private            10.503     16.583   0.633    0.528
## SCIM                             -3.080      5.348  -0.576    0.566
##
## (Dispersion parameter for gaussian family taken to be 2792030)
##
##     Null deviance: 3126422605  on 148  degrees of freedom
## Residual deviance:  345330911  on 128  degrees of freedom
##   (22 observations deleted due to missingness)
## AIC: 2650.6
##
## Number of Fisher Scoring iterations: 25
```

```
summary(m8_sev)
```

```
##
## Call:
## glm(formula = mean_ergo ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_ergo > 0))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -13581.8   -3107.3    -523.2    2966.3   24416.3
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.6040864  0.9084908   9.471  < 2e-16 ***
## sex2                           -0.0263679  0.0672991  -0.392   0.6955
## age_group]31,45]                0.0495742  0.0979823   0.506   0.6132
## age_group]45,60]                0.0500812  0.0865163   0.579   0.5631
## age_group]60,75]               -0.0197424  0.0923340  -0.214   0.8308
## age_group>75                   -0.2040179  0.1237774  -1.648   0.1003
## group1C5-C8, ASIA A,B,C,D       0.0003473  0.0643067   0.005   0.9957
## group1Paraplegia, ASIA A,B,C,D -0.5660384  0.0735236  -7.699 1.82e-13 ***
## etiologySickness               -0.1552683  0.0691305  -2.246   0.0254 *
## main_cost2                      0.2377535  0.4448750   0.534   0.5934
## main_cost3                     -0.7445246  0.8070542  -0.923   0.3570
## main_cost4                      0.0710863  0.0785101   0.905   0.3659
## main_cost5                      0.2579424  0.3449024   0.748   0.4551
## stay_aft_leav1                  1.4290166  0.9017172   1.585   0.1140
## stay_aft_leav4                  0.7004204  1.1240110   0.623   0.5336
## stay_aft_leav5                  0.6933099  0.9363476   0.740   0.4596
## stay_aft_leav6                  1.1965603  0.9178090   1.304   0.1933
## stay_aft_leavNurse              1.3037582  0.9030103   1.444   0.1498
```

```
## ins_classPrivate                0.0010054  0.0869549   0.012   0.9908
## ins_classSemi-Private          -0.0792856  0.0953683  -0.831   0.4064
## SCIM                           -0.0137589  0.0017700  -7.774 1.11e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 26623995)
##
##     Null deviance: 1.6088e+10  on 333  degrees of freedom
## Residual deviance: 8.3333e+09  on 313  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 6680.7
##
## Number of Fisher Scoring iterations: 8
```

```
summary(m9_sev)
```

```
##
## Call:
## glm(formula = mean_non_med ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_non_med > 0))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6227.9  -1904.1   -330.5   1662.5  10967.9
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.307783   0.639178  12.998  < 2e-16 ***
## sex2                           -0.055464   0.058219  -0.953  0.34148
## age_group]31,45]                0.057481   0.093394   0.615  0.53869
## age_group]45,60]                0.224971   0.078270   2.874  0.00433 **
## age_group]60,75]                0.062822   0.084594   0.743  0.45826
## age_group>75                   -0.137635   0.111710  -1.232  0.21884
## group1C5-C8, ASIA A,B,C,D      -0.148267   0.062480  -2.373  0.01825 *
## group1Paraplegia, ASIA A,B,C,D -0.258031   0.059666  -4.325 2.05e-05 ***
## etiologySickness               -0.036213   0.058626  -0.618  0.53723
## main_cost2                     -0.410457   0.588289  -0.698  0.48587
## main_cost3                     -0.074198   0.391923  -0.189  0.84997
## main_cost4                      0.051389   0.070133   0.733  0.46427
## main_cost5                     -0.063088   0.376382  -0.168  0.86699
## stay_aft_leav1                  1.180345   0.631131   1.870  0.06239 .
## stay_aft_leav4                  0.582496   0.789351   0.738  0.46110
## stay_aft_leav5                  0.591795   0.669576   0.884  0.37746
## stay_aft_leav6                  0.870858   0.652978   1.334  0.18328
## stay_aft_leavNurse              1.120197   0.632274   1.772  0.07742 .
## ins_classPrivate               -0.154739   0.085494  -1.810  0.07126 .
## ins_classSemi-Private          -0.162328   0.087696  -1.851  0.06510 .
## SCIM                           -0.015379   0.001582  -9.718  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 8582108)
```

```
##
##     Null deviance: 5035656764  on 334  degrees of freedom
## Residual deviance: 2694773432  on 314  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 6321.3
##
## Number of Fisher Scoring iterations: 7
```

```
summary(m10_sev)
```

```
##
## Call:
## glm(formula = mean_physt ~ sex + age_group + group1 + etiology +
##     main_cost + stay_aft_leav + ins_class + SCIM, family = gaussian(link = "log"),
##     data = filter(data_regression, mean_physt > 0))
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -24254   -5445    -861    4599   51375
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     8.622425   1.142184   7.549 4.73e-13 ***
## sex2                            0.068500   0.063476   1.079 0.281347
## age_group]31,45]                0.042310   0.096880   0.437 0.662605
## age_group]45,60]                0.154530   0.084367   1.832 0.067947 .
## age_group]60,75]                0.015630   0.091745   0.170 0.864830
## age_group>75                   -0.226761   0.137178  -1.653 0.099315 .
## group1C5-C8, ASIA A,B,C,D      -0.001012   0.074064  -0.014 0.989111
## group1Paraplegia, ASIA A,B,C,D -0.268937   0.071340  -3.770 0.000195 ***
## etiologySickness               -0.047562   0.072591  -0.655 0.512813
## main_cost2                      0.184147   0.337927   0.545 0.586184
## main_cost3                     -0.888515   1.187989  -0.748 0.455067
## main_cost4                      0.161070   0.080293   2.006 0.045704 *
## main_cost5                      0.069926   0.362191   0.193 0.847034
## stay_aft_leav1                  1.620283   1.136498   1.426 0.154948
## stay_aft_leav4                  0.738261   1.327116   0.556 0.578406
## stay_aft_leav5                  0.639461   1.187187   0.539 0.590518
## stay_aft_leav6                  1.171180   1.155053   1.014 0.311377
## stay_aft_leavNurse              1.368379   1.138374   1.202 0.230246
## ins_classPrivate                0.028121   0.083882   0.335 0.737666
## ins_classSemi-Private          -0.156763   0.099471  -1.576 0.116034
## SCIM                           -0.009026   0.001571  -5.745 2.16e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 80491598)
##
##     Null deviance: 3.7726e+10  on 336  degrees of freedom
## Residual deviance: 2.5435e+10  on 316  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 7113.3
##
## Number of Fisher Scoring iterations: 8
```

```
## R squared Mc Fadden
with(summary(m1_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.5275237
```

```
with(summary(m2_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.1423553
```

```
with(summary(m3_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.1985512
```

```
with(summary(m4_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.3234015
```

```
with(summary(m5_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.6346521
```

```
with(summary(m6_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.3216839
```

```
with(summary(m7_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.8895444
```

```
with(summary(m8_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.4820339
```

```
with(summary(m9_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.4648616
```

```
with(summary(m10_sev), 1 - deviance/null.deviance)
```

```
## [1] 0.325797
```

Bibliography

Bladt, Martin, Michel Fuino, Aleksandr Shemendyuk, and Joël Wagner. 2023. "Modelling the Burden of Long-Term Care for Institutionalised Elderly Based on Care Duration and Intensity." *Annals of Actuarial Science* 17 (1): 83–117.

Frees, Edward W, Jie Gao, and Marjorie A Rosenberg. 2011. "Predicting the Frequency and Amount of Health Care Expenditures." *North American Actuarial Journal* 15 (3): 377–92.

Panjer, Harry H, and Gordon E Willmot. 1983. "Compound Poisson Models in Actuarial Risk Theory." *Journal of Econometrics* 23 (1): 63–76.

Schober, Patrick, and Thomas R Vetter. 2021. "Count Data in Medical Research: Poisson Regression and Negative Binomial Regression." *Anesthesia & Analgesia* 132 (5): 1378–79.

Yang, Shengping, and Gilbert Berdine. 2015. "The Negative Binomial Regression." *The Southwest Respiratory and Critical Care Chronicles* 3 (10): 50–54.

Yee, Thomas W. 2015. *Vector Generalized Linear and Additive Models: With an Implementation in r*. Vol. 10. Springer.