Medical Care - Zero-Inflated and Zero-Hurdle-Model

November 1, 2024

First the medcare data are loaded:

```
> library(catdata)
> data(medcare)
> attach(medcare)
```

The dependent variable "ofp" (numbers of physician visits) is a count variable, so a poisson-family glm seems to be a good choice.

```
> med1=glm(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school,
+          family=poisson,data=medcare[male==1 & ofp<=30,])
> summary(med1)
```

In many real-world datasets the variance of count-data is higher than predicted by the Poisson distribution, so we fit a quasi-Poisson model with dispersion parameter.

```
> med2=glm(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school,
+          family=quasipoisson,data=medcare[male==1 & ofp<=30,])
> summary(med2)
```

With an estimated dispersion parameter of 4.69 the standard errors are much bigger now. An alternative to a quasi-poisson model is to use the negative binomial distribution.

```
> library(MASS)
> med3=glm.nb(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school,
+             data=medcare[male==1 & ofp<=30,])
> summary(med3)
```

In this model the standard errors are slightly lower with the result that "healthexcellent" and "married" are now significant. (level=0.05) In count data there are often much more zeros than expected. Therefore one can fit a "zero-inflated" model using the pscl package. In the first "zero-inflated" model one assumes that the occurence of zeros does depend on covariates:

```
> library(pscl)
```

```
> med4=zeroinfl(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school|1,
+               data=medcare[male==1 & ofp<=30,])
> summary(med4)
```

In the second "zero-inflated" model the occurence of zeros can depend on covariates:

```
> med5=zeroinfl(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school,
+               data=medcare[male==1 & ofp<=30,])
> summary(med5)
```

An alternative to "zero-inflation" is the "zero-hurdle" model. In the following similar models as above are fitted.

```
> med6=hurdle(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school|1
+             ,data=medcare[male==1 & ofp<=30,])
> summary(med6)
```

```
> med7=hurdle(ofp ~ hosp+healthpoor+healthexcellent+numchron+age+married+school,
+             data=medcare[male==1 & ofp<=30,])
> summary(med7)
```