

## Zero-truncated negative binomial regression

- ▶ Zero-truncated negative binomial regression is used to model count data for which the value zero cannot occur and for which over dispersion exists.

## Zero-truncated negative binomial regression

- ▶ To fit the zero-truncated negative binomial model, we use the `vglm` function in the VGAM package.
- ▶ This function fits a very flexible class of models called vector generalized linear models to a wide range of assumed distributions.
- ▶ In our case, we believe the data come from the negative binomial distribution, but without zeros.
- ▶ Thus the values are strictly positive poisson, for which we use the positive negative binomial family via the `posnegbinomial` function passed to `vglm`.

# Zero-truncated negative binomial regression

## Fitting the Model with R

We will use the *hospitalstay* data again.

```
m1 <- vglm(stay ~ age + hmo + died,  
  family = posnegbinomial(),  
  data = hospitalstay)
```

# Zero Truncated Negative Binomial Regression

```
summary(m1)
##
## Call:
## vglm(formula = stay ~ age + hmo + died,
##       family = posnegbinomial(),
##       data = hospitalstay)
##
## Pearson Residuals:
##           Min      1Q  Median      3Q      Max
## log(munb)  -1.4 -0.70  -0.23  0.45  9.8
## log(size) -14.1 -0.27   0.45  0.76  1.0
```

# Zero Truncated Negative Binomial Regression

## Coefficients:

##		Estimate	Std. Error	z	value
##	(Intercept):1	2.408	0.072	33.6	
##	(Intercept):2	0.569	0.055	10.4	
##	age	-0.016	0.013	-1.2	
##	hmo1	-0.147	0.059	-2.5	
##	died1	-0.218	0.046	-4.7	

# Zero Truncated Negative Binomial Regression

- ▶
- ▶ The first intercept is what we know as the typical intercept.
- ▶ The second is the **over dispersion parameter**,  $\alpha$ .
- ▶ The number of linear predictors is two, one for the expected mean  $\lambda$  and one for the over dispersion.
- ▶ Next the dispersion parameter is printed, assumed to be one after accounting for overdispersion.

## Zero Truncated Negative Binomial Regression

- ▶ The value of the coefficient for age,  $-0.0157$  suggests that the log count of stay decreases by  $0.0157$  for each year increase in age.
- ▶ The coefficient for hmo,  $-0.1471$  indicates that the log count of stay for HMO patient is  $0.1471$  less than for non-HMO patients.
- ▶ The log count of stay for patients who died while in the hospital was  $0.2178$  less than those patients who did not die.

# Zero Truncated Negative Binomial Regression

- ▶ The value of the constant 2.4083 is the log count of the stay when all of the predictors equal zero.
- ▶ The value of the second intercept, the over dispersion parameter,  $\alpha$  is 0.5686.
- ▶ To test whether we need to estimate over dispersion, we could fit a zero-truncated Poisson model and compare the two.  
(Not Covered).