Intro to Poisson Regression

When, and when not, to use model selection

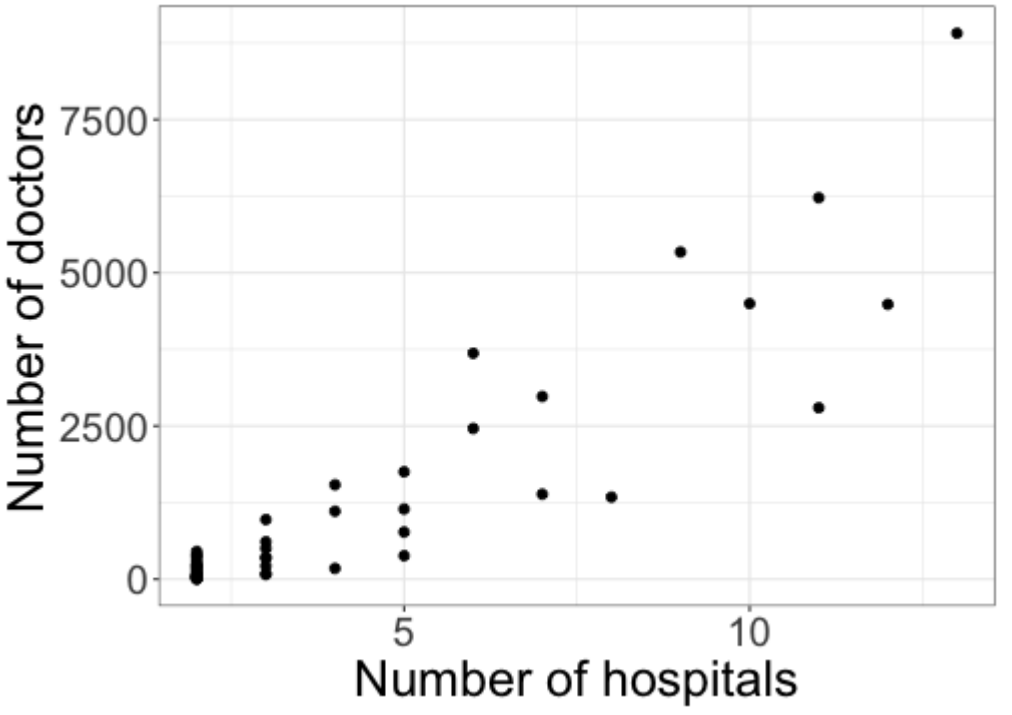
Count variables

Data: Data on medical facilities and doctors from a sample of 53 different counties in the US. Variables include:

- `MDs`: the number of medical doctors in the county
- `Hospitals`: the number of hospitals in the county

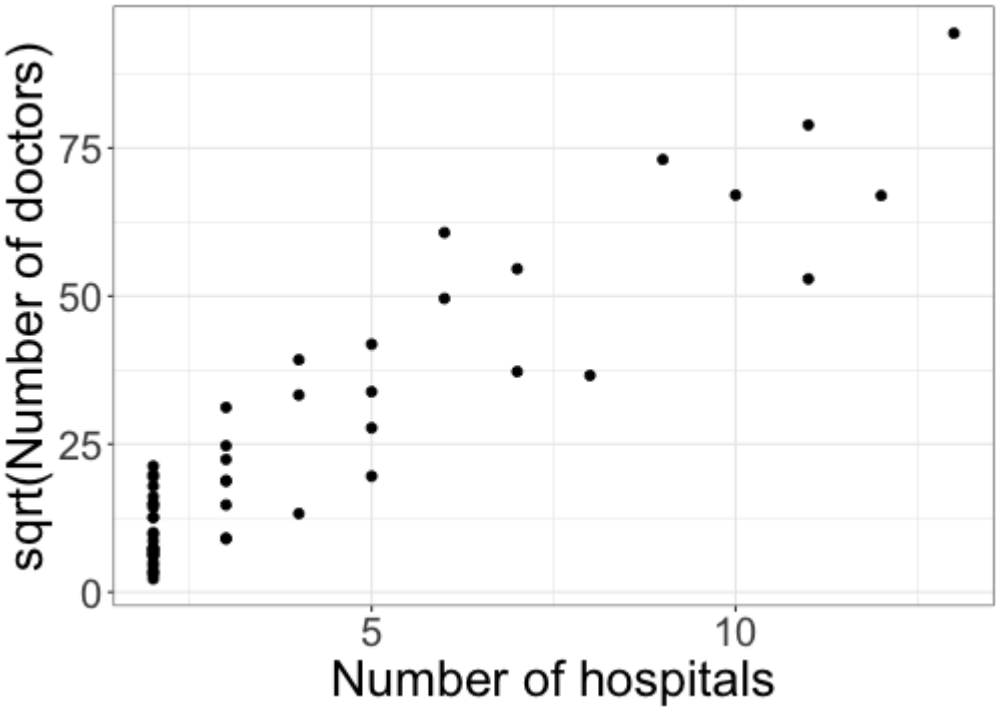
Research question: Can we model the relationship between the number of hospitals and the number of doctors?

Plotting the data



Does a linear regression model seem appropriate for this relationship?

Trying a transformation



Is a linear regression model appropriate now?

Poisson regression

Fitting the Poisson regression model

```
m1 <- glm(MDs ~ Hospitals, data = CountyHealth,
          family = poisson)
summary(m1)
```

```
...
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.116896   0.009801   522.1   <2e-16 ***
## Hospitals   0.312442   0.001048   298.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 111627  on 52  degrees of freedom
## Residual deviance:  22799  on 51  degrees of freedom
## AIC: 23197
...
```

Interpreting the Poisson regression model

```
m1 <- glm(MDs ~ Hospitals, data = CountyHealth,
          family = poisson)
summary(m1)
```

```
...
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.116896   0.009801   522.1   <2e-16 ***
## Hospitals   0.312442   0.001048   298.2   <2e-16 ***
...
```

Exponential dispersion models