Lecture 7

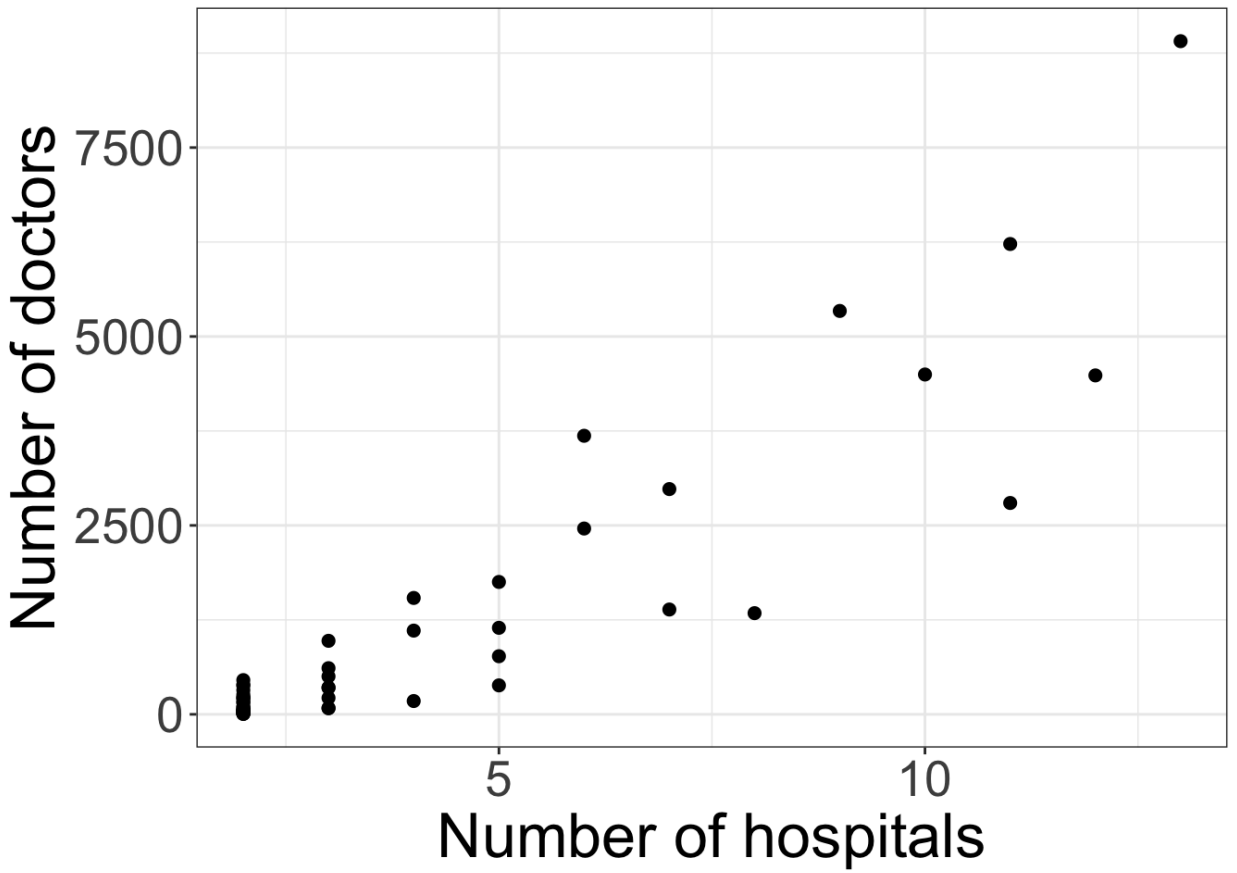
Count variables

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

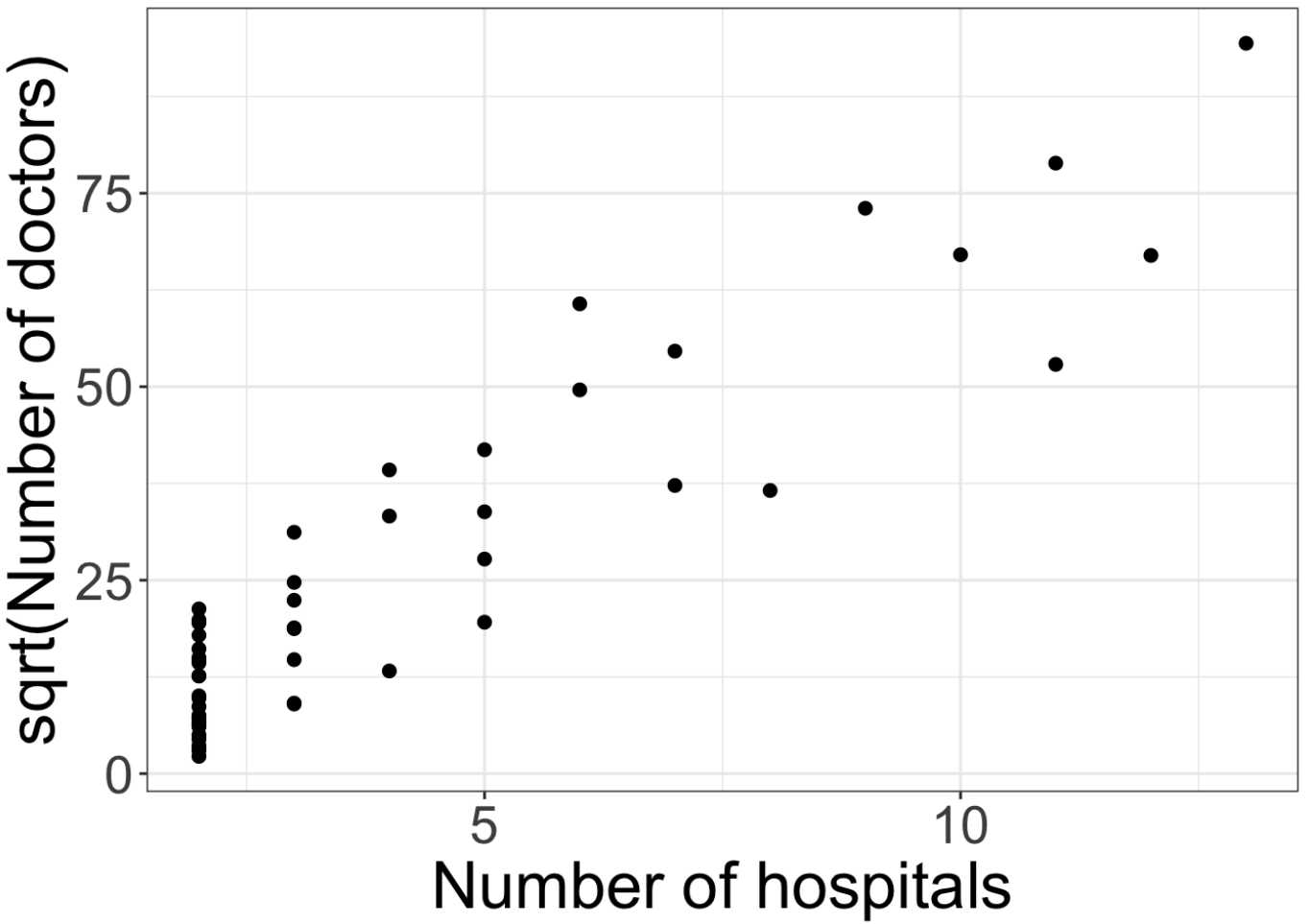
Research question: Can we model the relationship between the number of hospitals and the number of doctors?

Plotting the data



Question: 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

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

```
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 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

Number of Fisher Scoring iterations: 5
...
```

Interpreting the Poisson regression model

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

```
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
...
```

Exponential dispersion models