doctors

Chris Parrish

January 7, 2016

doctors

references:
- Cannon, et al., Stat2, chapter 01, example 1.6

Import the data.

```
data <- read.csv("MetroHealth83.csv", header=TRUE)
head(data[ , c(1, 2, 4)])
```

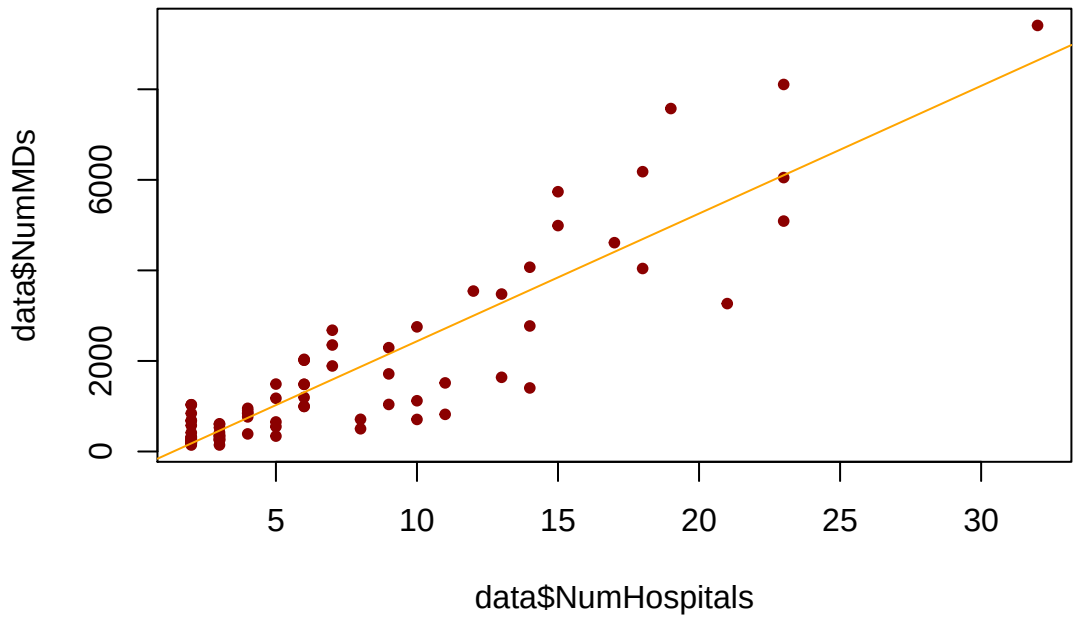
```
##                             City NumMDs NumHospitals
## 1       Holland-Grand Haven, MI    349            3
## 2             Louisville, KY-IN   4042           18
## 3              Battle Creek, MI    256            3
## 4                   Madison, WI   2679            7
## 5             Fort Smith, AR-OK    502            8
## 6 Sarasota-Bradenton-Venice, FL   2352            7
```

```
dim(data)
```

```
## [1] 83 16
```

View the data.

```
plot(data$NumHospitals, data$NumMDs,
     pch=20, col="darkred")
doctors.lm <- lm(NumMDs ~ NumHospitals, data=data)
abline(doctors.lm, col="orange")
```



Linear model.

$\widehat{NumMDs}$ = -385.102 + 282.009 $NumHospitals$

```
options(show.signif.stars=FALSE)
summary(doctors.lm)
```

```
##
## Call:
## lm(formula = NumMDs ~ NumHospitals, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2270.09  -263.44    58.08   309.02  2601.93
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -385.10     138.26  -2.785  0.00666
## NumHospitals   282.01      14.42  19.563  < 2e-16
##
## Residual standard error: 833.2 on 81 degrees of freedom
## Multiple R-squared:  0.8253, Adjusted R-squared:  0.8232
## F-statistic: 382.7 on 1 and 81 DF,  p-value: < 2.2e-16
```

```
anova(doctors.lm)
```

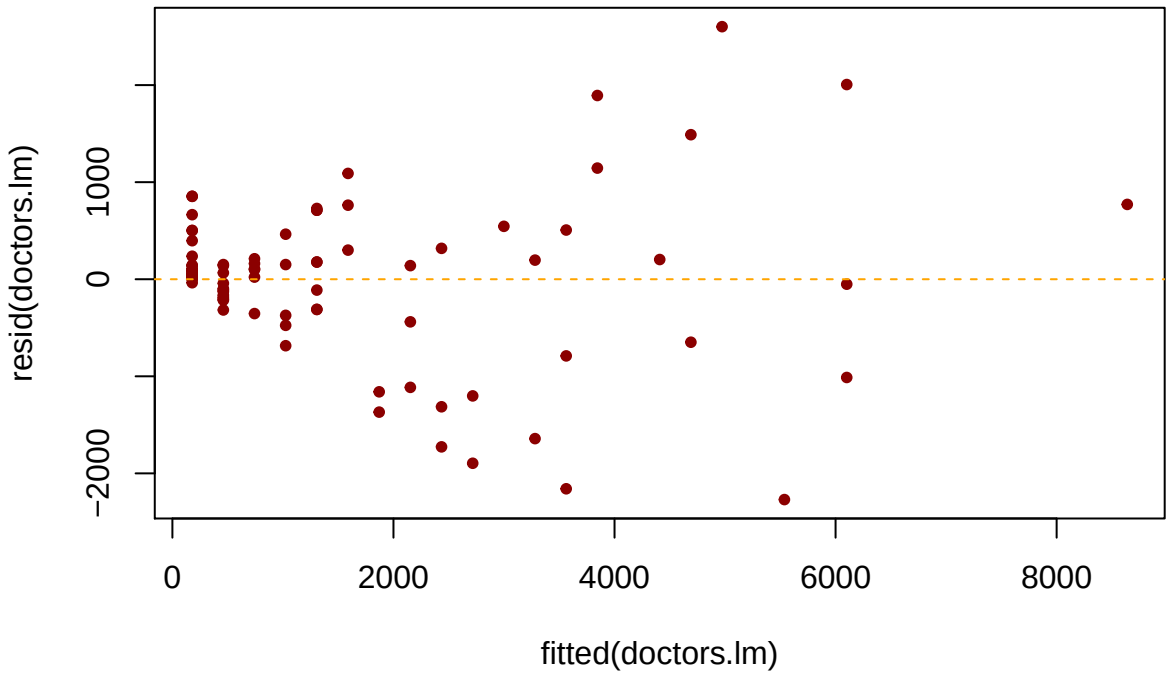
```
## Analysis of Variance Table
##
## Response: NumMDs
##              Df    Sum Sq   Mean Sq F value    Pr(>F)
## NumHospitals  1 265700224 265700224  382.69 < 2.2e-16
## Residuals    81  56237663    694292
```

Regression (= residual) standard error.

$\widehat{\sigma_e} = \sqrt{MSE}$ = 833.242

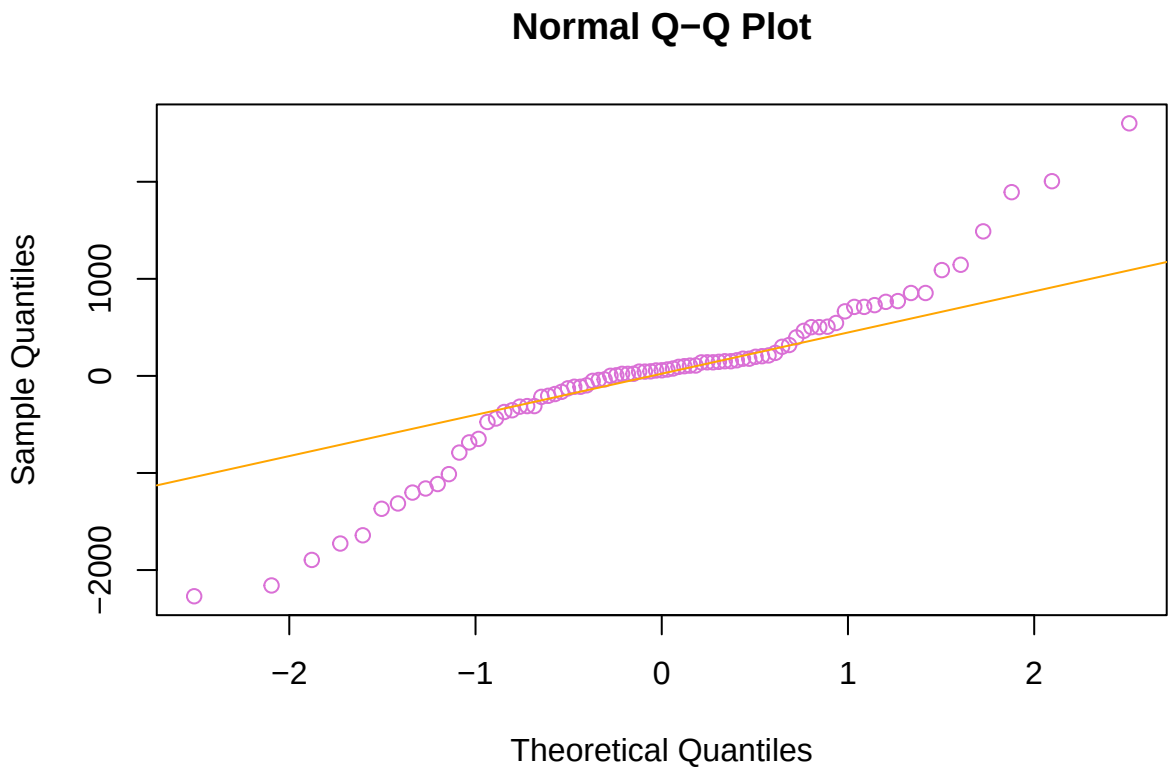
Residuals.

```
plot(fitted(doctors.lm), resid(doctors.lm),
     pch=20, col="darkred")
abline(h=0, col="orange", lty="dashed")
```



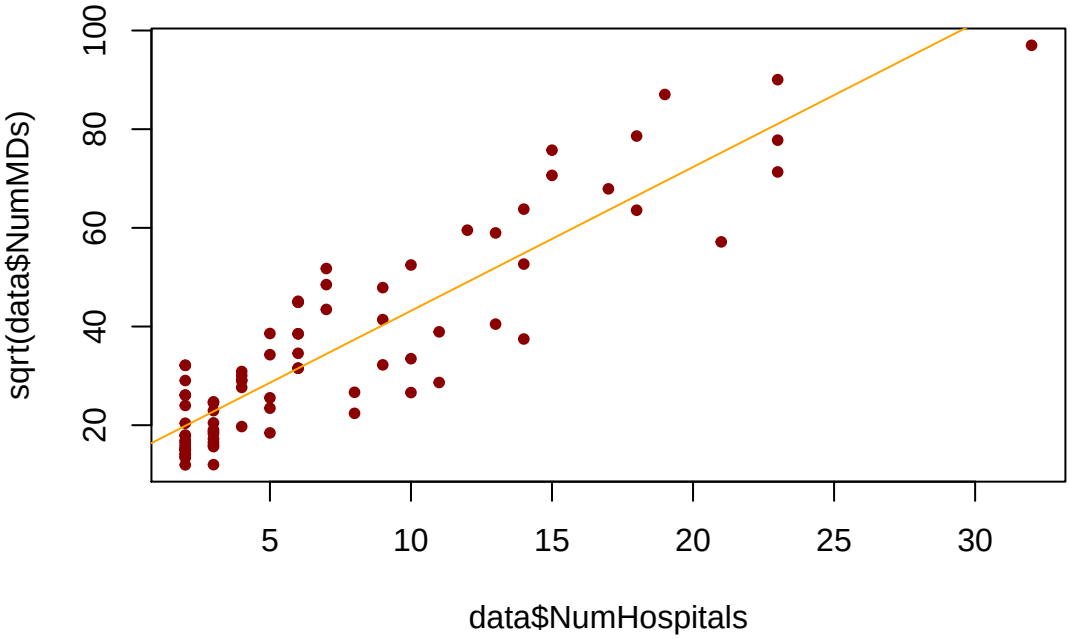

```
hist(resid(doctors.lm), col="wheat")
```


```
qqnorm(resid(doctors.lm),
       col="orchid")
qqline(resid(doctors.lm), col="orange")
```



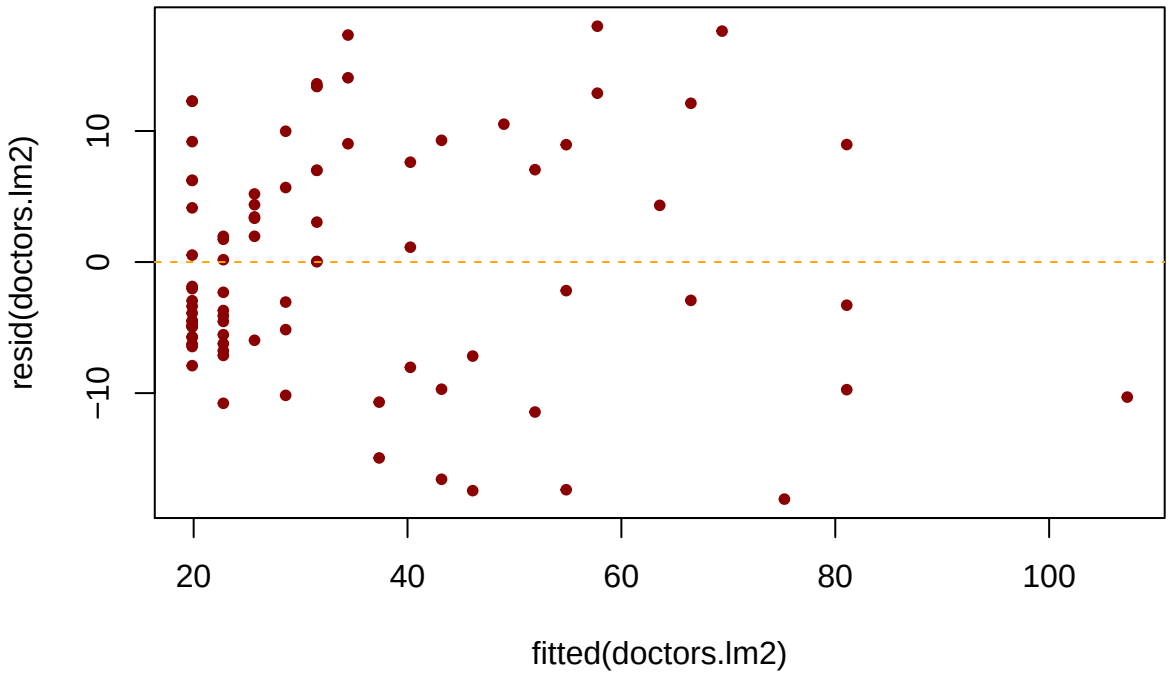
Transformation.

```
plot(data$NumHospitals, sqrt(data$NumMDs),
     pch=20, col="darkred")
doctors.lm2 <- lm(sqrt(NumMDs) ~ NumHospitals, data=data)
abline(doctors.lm2, col="orange")
```



Residuals.

```r
plot(fitted(doctors.lm2), resid(doctors.lm2),
     pch=20, col="darkred")
abline(h=0, col="orange", lty="dashed")
```

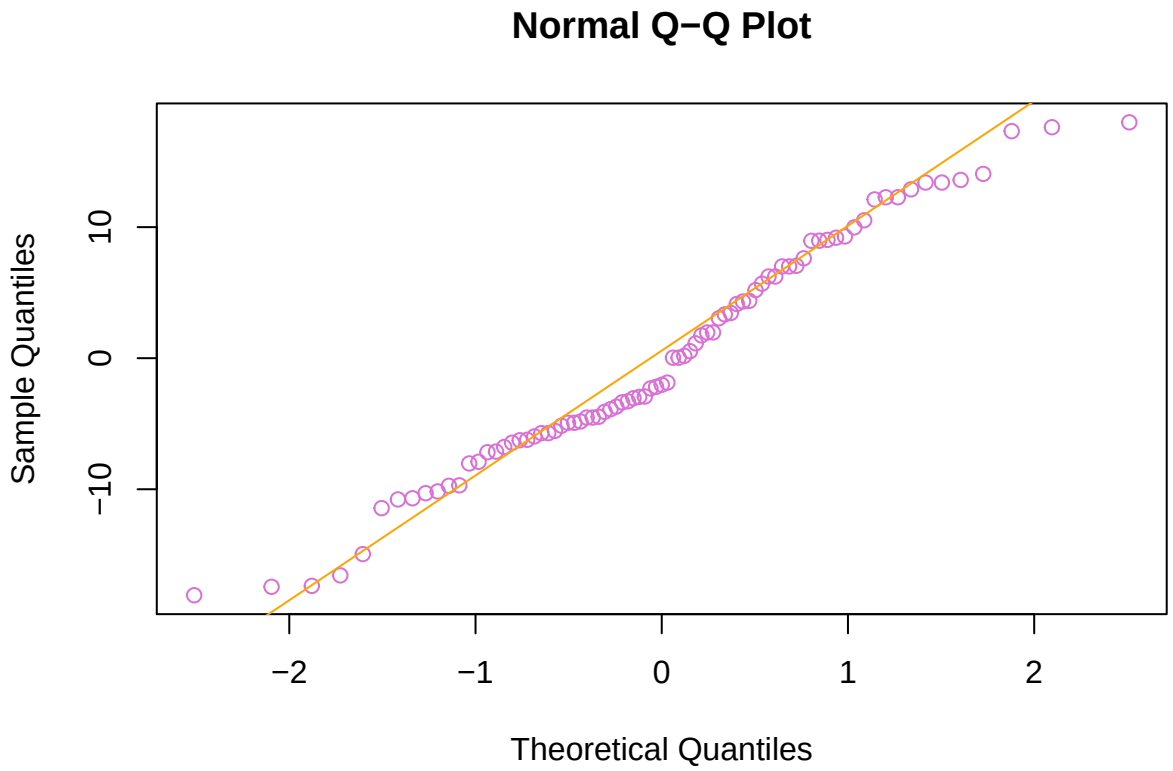



```r
doctors.lm2
```

```
##
## Call:
## lm(formula = sqrt(NumMDs) ~ NumHospitals, data = data)
##
## Coefficients:
##  (Intercept)  NumHospitals
##       14.033         2.915
```

```r
hist(resid(doctors.lm2), col="wheat")
```


```
qqnorm(resid(doctors.lm2),
       col="orchid")
qqline(resid(doctors.lm2), col="orange")
```



Prediction.

```
new.data <- data.frame(NumHospitals=18)
y.hat <- predict(doctors.lm2, new.data)
NumMDs.hat <- y.hat^2
NumMDs.hat
```

```
##        1
## 4422.201
```

New linear model.

$\widehat{NumMDs} = (-385.102 + 282.009 \cdot NumHospitals)^2$

Illustration.

```
y.hat <- function(x){
  a <- 14.033
  b <- 2.915
  y.hat <- (a + b * x)^2
  return(y.hat)
}
plot(data$NumHospitals, data$NumMDs,
     pch=20, col="darkred")
curve(y.hat, from=2, to=30,
      col="olivedrab", add=TRUE)
```