Linear Regression Notes

Chirag Lakhani

May 23, 2015

We will use the HERS dataset, which is a clinical trial that was used to look at the effects of hormone-therapy on 2,764 post-menopausal women. For more information see http://pages.pomona.edu/~jsh04747/courses/math150/HERS.pdf

---

setwd("~/Documents/gitProjects/notes/")

Simple Linear Regression

If one is given a set of points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathbb{R}^2$ where

- y - response variable
- x - predictor variable

then linear regression is an approach to modeling the linear relationship between the response and predictor variable. The model is of the form below:

$$y = \beta_0 + \beta_1 x$$

The goal will be to calculate the coefficients β_0, β_1 that best fit the data.

```
library(ggplot2)
hersdata <- read.delim("hersdata.txt")
head(hersdata[,c('BMI','LDL')])
```

```
##      BMI   LDL
## 1 23.69 122.4
## 2 28.62 241.6
## 3 42.51 166.2
## 4 24.39 116.2
## 5 21.90 150.6
## 6 29.05 137.8
```