Stats 401 Lab 4

Naomi Giertych

9/28/2018

Announcements

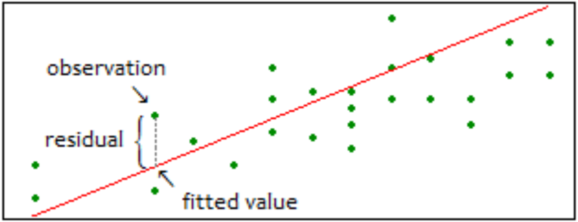
- Homework 3 is due today
- Quiz 1 is on October 5 (next week!)
- In lab
- Approximately 40 minutes
- Let us know NOW if you have special accomodations

Quiz Topics

- Summations
- R Exercises
- Basic matrix computations
- Fitting a linear model
- Essentially HW 1-4 and notes 1-4 (ending Wednesday)

Examining the Linear Model

- Recall the sample linear model in matrix form: $\mathbf{y} = \mathbb{X}\mathbf{b} + \mathbf{e}$
- The goal is to fit a curve that best fits the observed data
- We do this by minimizing the residuals



Minimizing the Residuals

- Residual Sum of Squares (RSS):

$$\sum_{i=1}^{n} (y_i - mx_i - c)^2$$

where $(y_i - mx_i - c)$ is the residual of observation i

- Recall that we find the minimum and maximum of a function by taking the derivative and setting it equal to 0.
- Since RSS depends on m and c, we need to solve $\partial RSS/\partial m = 0$ and $\partial RSS/\partial c = 0$
- Note: $RSS \geq 0$ and arbitrarily large for poor choices of m and c, it has a minimum but not a maximum.

Minimizing the Residuals, cont.

- The **general** solution to these equations is precisely: $\mathbf{b} = [\mathbb{X}^{\mathbb{T}}\mathbb{X}]^{-1}\mathbb{X}^{T}\mathbf{y}$

(You will NOT be required to reproduce these results.)

- Constructing the **general RSS**:
- residual for unit i is $e_i = y_i - [\mathbb{X}\mathbf{b}]_i$
- $RSS = \sum_{i=1}^{n} e_i^2$

Probability

- Recall: **random variable** X is a random number with probabilities assigned to the outcomes
- Recall: A random variable can take on discrete (e.g. a die: $\{1, 2, 3, 4, 5, 6\}$) or continuous values (e.g. weight following a normal distribution)

- Suggestion: Review from STATS 250 the concepts of expected value and variance and the properties of common distrbutions such as the normal.

In Lab Exercises (Part 1)

- Show $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2) - n(\bar{x})^2$.
- Show how $\mathbf{y} = \mathbb{X}\mathbf{b}$, with n observations, can be written as a sum. $(\mathbf{y}_{n\times 1}, \mathbb{X}_{n\times p}, \mathbf{b}_{p\times 1})$
- Show how $\sum_{i=1}^{n} 3x_i$ can be written in matrix form.

In Lab Exercises (Part 2)

- Let Y be a discrete random variable that takes on values 0, 1, and 2 with probabilities 0.5, 0.3, and 0.2 respectively.
- What is the expected value of Y?
- What is the variance of Y?
- (Challenge): Suppose instead that Y is a continuous random variable from [0,3]. What would be a natural extension of the calculation of the expected value of Y; i.e. how would you sum across [0,3]? (No calculations necessary.)

In Lab Exercises (Part 3)

- Suppose we define $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$ as follows,

```
A
```

```
##      [,1] [,2]
## [1,]    0    3
## [2,]    1    2
## [3,]   -2   -2
```

```
B
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]   -2    1
```

In Lab Exercises (Part 2) (cont.)

```
C
```

```
##      [,1] [,2]
## [1,]    0    1
## [2,]    0    1
## [3,]    0    1
## [4,]    0    1
```

Write the commands that would produce these matrices in R.

Lab Ticket

Using the matrices from the lab exercise, calculate the matrices returned by following r commands:

1. A %*% B
2. t(A)
3. solve(B)

- Sugar pumpkins are coming into season! They're often used to make pumpkin pie. Let X be the weight of a sugar pie pumpkin (in lbs) and suppose it is known that X~normal(2.5, 0.5). Using `pnorm()`, find the probability that a pumpkin chosen at random will weigh less than 1.9 lbs.
- (Challenge Question) Suppose we draw 20 sugar pumpkins at random. What is the probability that the average weight of the pumpkins will be less than 2.3 lbs. (Hint: It may be useful to review the Central Limit Theorem from STATS 250.)