Stats 401 Lab 4

Naomi Giertych

9/28/2018

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 1) Solutions

- Show $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2) - n(\bar{x})^2$.
  - $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x})$
  - $= \sum_{i=1}^{n}(x_i^2) + 2\bar{x}\sum_{i=1}^{n}(x_i) + \bar{x}\sum_{i=1}^{n}(1)$
  - however $\sum_{i=1}^{n}(x_i) = n(\frac{1}{n}\sum_{i=1}^{n}(x_i)) = n\bar{x}$
  - however $\bar{x}\sum_{i=1}^{n}(1) = n\bar{x}$
  - $= \sum_{i=1}^{n}(x_i^2) + 2\bar{x}\sum_{i=1}^{n}(x_i) + \bar{x}\sum_{i=1}^{n}(1) = \sum_{i=1}^{n}(x_i^2) + 2n\bar{x}^2 - n\bar{x}^2$
  - $= \sum_{i=1}^{n}(x_i^2) - n(\bar{x})^2$

In Lab Exercises (Part 1) Solutions cont.

- Show how $\mathbf{y} = \mathbb{X}\mathbf{b}$, with n observations, can be written as a sum. ($\mathbf{y}_{n\times 1}, \mathbb{X}_{n\times p}$, $\mathbf{b}_{p\times 1}$)
  - $$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$
  - $$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11}b_1 & x_{12}b_2 & \dots & x_{1p}b_p \\ x_{21}b_1 & x_{22}b_2 & \dots & x_{2p}b_p \\ \vdots & & & \\ x_{n1}b_1 & x_{n2}b_2 & \dots & x_{np}b_p \end{bmatrix}$$
  - $$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} x_{1j}b_j \\ \sum_{j=1}^{n} x_{2j}b_j \\ \vdots \\ \sum_{j=1}^{n} x_{nj}b_j \end{bmatrix}$$

In Lab Exercises (Part 1) Solutions cont.

- Show how $\sum_{i=1}^{n} 3x_i$ can be written in matrix form.
  - $\sum_{i=1}^{n} 3x_i = 3x_1 + 3x_2 + \cdots + 3x_n$
  - $3x_1 + 3x_2 + \cdots + 3x_n = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$
  - $\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ \vdots \\ 3 \end{bmatrix} = \mathbf{x3}$

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 2) Solutions

- $E(Y) = 0 \times (0.5) + 1 \times (0.3) + 2 \times (0.2) = 0.7$
- $Var(Y) = E[(Y - E(Y))^2] = E[Y^2 - 2YE(Y) + E(Y)^2]$
  - $E[Y^2 - 2YE(Y) - E(Y^2)] = E(Y^2) - 2E(Y)E(Y) + E(Y)^2$
  - $E(Y^2) - 2E(Y)E(Y) + E(Y)^2 = E(Y^2) - E(Y)^2$
- $E(Y^2) = 0^2 \times (0.5) + 1^2 \times (0.3) + 2^2 \times (0.2) = 1.1$
- $Var(Y) = 1.1 - (0.7)^2 = 1.1 - 0.49 = 0.61$
- Challenge solution: The natural extension of the summation over discrete values of $\{0, 1, 2, 3\}$ would be the integral from 0 to 3 (across [0, 3]).

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.

In Lab Exercises (Part 3) Solutions.

- $\mathbb{A}$: `matrix(c(0, 1, -2, 3, 2, -2), nrow = 3)`
- $\mathbb{B}$: `matrix(c(1, -2, 0, 1), nrow = 2)`
- $\mathbb{C}$: `matrix(c(rep(0, 4), rep(1, 4)), nrow = 4)`

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.)

Lab Ticket Solutions

1. A %*% B
   - $\begin{bmatrix} 0 & 3 \\ 1 & 2 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -6 & 3 \\ -3 & 2 \\ 2 & -2 \end{bmatrix}$
2. t(A)
   - $\begin{bmatrix} 0 & 1 & -2 \\ 3 & 2 & -2 \end{bmatrix}$
3. solve(B)
   - $\frac{1}{1} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$

Lab Ticket Solutions cont.

- Using `pnorm()`, find the probability that a pumpkin chosen at random will weigh less than 1.9 lbs.
  - `pnorm(1.9, 2.5, 0.5)` = 0.1150697
- (Challenge Question)
  - Since $X \sim N(2.5, 0.5)$, from the CLT we know that $\bar{X} \sim N(2.5, \frac{0.5}{\sqrt{20}})$
  - Then, the probability that the average weight of the pumpkins will be less than 2.3 lbs is `pnorm(2.3, 2.5, 0.5/sqrt(20))` = 0.03681914