Stats 401 Lab 2

Sanjana Gupta

9/7/2018

Office hours

- Dr. Ionides. Mon 10-11am, Wed 3-4pm (his office)

At 2165 USB, the Science Learning Center Annex

- Sanjana Gupta. Mon 4:30-5:30pm, Tu 11:30am-12:30pm
- Ed Wu. Tu 12:30-2:30pm
- Naomi Giertych. Thu 9-11 am

Homework

- Out of 10 points
- **0 points** if there is no statement of sources
- Provide the code if the question requests

Swirl tutorial

We finished lessons 1/3/4 in HW1.

- Any techical difficulties encountered working with swirl?
- Any questions about materials introduced in the tutorial?

Swirl tutorial

You are asked to complete lessons 5/6/7/9 for HW2. Lesson 9 can be a little bit harder.

- We can go through parts of it together at the end of this lab (if we have time).
- You can always go to our office hours for help.

Topics covered in today's Lab

- R functions
  - R help: '?'
  - Apply function: 'apply()'
- Summation notation

R functions: help

- Access the documentation of functions by typing '?*name of function*'.
- Try '?mean', '?median'.
- What if you don't know the name of the inbuilt function? Try '??*name of concept*.
- Try '??columnnames', '??dimension'

R functions: Apply

- Applies a given function to a vector or rows/ columns of a matrix.
- See documentation by typing '?apply'

Let's see an example: Find the average GPA and ACT scores of the students from the dataset used in the last lab.

```
# Load the Dataset
gpa = read.table("CH01PR19.txt", header = T)

# Recall the dataset
#head(gpa)

# Find the mean of the columns
apply(gpa,2,mean)
```

```
##       GPA       ACT
##   3.07405 24.72500
```

Summation

This is simply a compressed form of writing addition of many terms.

Given n constants $x_1, x_2, .., x_m, x_{m+1}, .., x_n$,

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_{n-1} + x_n$$

In general,

$$\sum_{i=m}^{n} x_i = x_m + x_{m+1} + ... + x_{n-1} + x_n$$

Summation: Examples

- $\sum_{i=1}^{6} i = 1 + 2 + 3 + 4 + 5 + 6 = 21$
- $\sum_{i=3}^{5} i^2 = 3^2 + 4^2 + 5^2 = 9 + 16 + 25 = 50$
- $\sum_{i=1}^{9} 1 = \underbrace{1 + 1 + ... + 1}_{9 times} = 9$
- If $x_1 = 11$, $x_2 = 22$, $x_3 = 21$, $x_4 = 12$,
  $\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 11 + 22 + 21 + 12 = 66$

Note: Comparing to basic summation formula on prev slide,
$x_i = i$ in example 1, $x_i = i^2$ in example 2, $x_i = 1$ in example 3

Summation: Basic Properties

For any given (fixed) numbers n,m and constants c,d

Basic addition

- $\sum_{i=1}^{n} 1 = \underbrace{1+1+....+1}_{ntimes} = n \times 1 = n$
- $\sum_{i=1}^{n} d = \underbrace{d+d+....+d}_{ntimes} = n \times d$
- $\sum_{i=m}^{n} 1 = \underbrace{1+1+....+1}_{n-m+1times} = (n-m+1) \times 1 = n-m+1$
- $\sum_{i=m}^{n} d = \underbrace{d+d+....+d}_{n-m+1times} = (n-m+1) \times d = (n-m+1)d$

Addition of summations

- $\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (x_i + y_i)$
- $\sum_{i=1}^{4} (i + i^2) = \sum_{i=1}^{4} i + \sum_{i=1}^{4} i^2 = 10 + 30 = 40$

Summation: Basic Properties (ctd)

Scalar multiplication

- $$\begin{aligned} c\left(\sum_{i=1}^{n} x_i\right) &= c(x_1 + x_2 + \ldots + x_{n-1} + x_n) \\ &= cx_1 + cx_2 + \ldots + cx_{n-1} + cx_n \\ &= \sum_{i=1}^{n} c(x_i) \end{aligned}$$
- $5\sum_{i=1}^{3} i = 5(1+2+3) = 5\times 1 + 5\times 2 + 5\times 3 = \sum_{i=1}^{3} 5i$

Summation: Relating to linear model

Recall LM1, LM2 from ch1 notes:
Suppose our data are $\{y_1, y_2, \ldots, y_n\}$ and on each unit $\{\text{i}\}$ we have $\{\text{p}\}$ explanatory variables $\{x_{i1}, x_{i2}, \ldots, x_{ip}\}$. A linear model is for $i = 1, 2, \ldots, n$

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip} + e_i \qquad \text{(LM1)}$$

which is equivalent to

$$y_i = \sum_{j=1}^{p} x_{ij} b_j + e_i \qquad \text{(LM2)}$$

In-lab Activity

- Find the median GPA and ACT scores of the students from the dataset used in lab1 (CH01PR19.txt)
- Express the mean of $x_1, x_2, \ldots, x_n$ in summation notation. *(Hint: Recall that* $mean(x_1, x_2, \ldots, x_n) = \frac{x_1+x_2+\cdots+x_n}{n}$ *)*

In-lab Activity Part 1

- Let us find the median GPA and ACT scores of the students.

```
# Find the median scores
apply(gpa,2,median)
```

```
##      GPA      ACT
##   3.0775 25.0000
```

In-lab Activity Part 2

- Express mean in terms of summation

$$
\begin{aligned}
mean(x_1, x_2, \ldots, x_n) &= \frac{x_1 + x_2 + \cdots + x_n}{n} \\
&= \sum_{i=1}^{n} (x_i)/n \\
&= \frac{1}{n} \sum_{i=1}^{n} x_i
\end{aligned}
$$

Exit ticket

- Load the unemployment dataset from https://ionides.github.io/401f18/01/unemployment.csv
- See the first few observations using the head() command
- Find the average unemployment rate for each month *(use apply and mean function)*
- Recall the definition of standard deviation. Find the inbuild function in R for standard deviation (using the help function '??')
- Using this, find the standard deviation of the unemployment rate for each month