

# STATS 531 Homework 7

Paul Zhang

## 0. Libraries and Data

```
library(doParallel)
```

## 1. Greatlakes Cluster Test

### Feedbacks

- The test code runs almost perfectly on the server. The only issue I have encountered so far is that the waiting time before my batch job starts can be long sometimes
- For the record, I am familiar with Linux, but I have not used any Linux clusters before.
- For some reason, the code sometimes runs significantly faster on my laptop, which seems suspicious. To investigate whether the issue is related to parallelization, I modified `test.R` to include the number of cores used. The modified file can be found in the Appendix.

Below are the runtimes for the test code on the server and my laptop.

### Greatlakes Cluster

```
[paulzh@gl-login3 hw7]$ cat test.csv
","user.self","sys.self","elapsed","user.child","sys.child","cores"
"time0",5.224,0.29,5.527,0,0,36
"time1",0.124,0.826,1.777,0.53,0.122,36
"time2",0.102,0.748,1.172,1.816,0.511,36
"time3",0.239,0.578,1.166,5.422,1.75,36
"time4",0.969,0.877,2.044,5.539,2.277,36
```

### Local Env

```
[paulzhang@ubuntu Homework7(main)]$ cat test.csv
","user.self","sys.self","elapsed","user.child","sys.child","cores"
"time0",1.753,0.189,1.942,0,0,32
"time1",0.108,0.501,1.311,0.273,0.1,32
"time2",0.089,0.571,1.178,5.383,1.495,32
"time3",0.089,0.662,1.286,5.538,2.067,32
"time4",0.617,1.215,2.419,5.433,6.726,32
```

## Stop Typing SSH Addresses Over and Over

Manually entering SSH login addresses each time is infuriating. Fortunately, this can be easily fixed by adding the following entry to `~/.ssh/config`:

```
Host gl
  HostName greatlakes.arc-ts.umich.edu
  User unixname
```

```
Host glxfer
  HostName greatlakes-xfer.arc-ts.umich.edu
  User username
```

After that, the following commands

```
ssh username@greatlakes.arc-ts.umich.edu
scp myfile username@greatlakes-xfer.arc-ts.umich.edu:mydir
```

can be replaced by

```
ssh gl
scp myfile glxfer:mydir
```

**2.**