

Biostat 561: Homework 8

Instructor: Amy Willis, Biostatistics, UW

November 16, 2017

Homework due Thursday 30 November, 2 p.m.

No office hours Tu 28th due to travel; special office hours Monday 27 November, 3pm

Link to Homework 8 submission: <https://classroom.github.com/a/JfDr7BQP>

Question 1: Running R scripts from the command line

- a) Read through `se_ex1.R`, the script to compare model-based and sandwich standard errors of the model

$$\begin{aligned} X_i &\overset{i.i.d.}{\sim} N(0, 1), \\ u_i &\overset{i.i.d.}{\sim} N(0, 1), \text{ independent of the } X_i, \\ Y_i | X_i, u_i &= \beta_0 + \beta_1 X_i + \epsilon_i, \text{ where} \\ \epsilon_i &= |X_i| u_i. \end{aligned}$$

Based on reading the script, which parameter are we interested in estimating? How do you know?

- b) Open up an R session (e.g. in RStudio) and make sure that the package `sandwich` is installed on your computer.
- c) Open up a terminal and run the command

```
Rscript se_ex1.R B=10 seed=547 n=5 truebeta=2
```

What standard errors do you find based on each method?

- d) Now run the script `call_ex1_once.sh` via the command line:

```
./call_ex1_once.sh B=10 seed=547 n=5 truebeta=2
```

You may need to change your permissions to execute the script with `chmod` (see the slides).¹

What standard errors do you find for each model now?

Question 2: Running R scripts on cox

Log on to cox and run

```
./call_ex1_once.sh B=10000 seed=547 n=500 truebeta=2
```

Approximately how long did it take?

¹I kept getting "bad interpreter" errors, which I solved using the advice at stackoverflow.com/questions/2920416/configure-bin-shm-bad-interpreter

Question 3: Running R scripts on the clusters

Time to use the cluster! Use `submit_ex1_batch.sh` and `submit_ex2_batch.sh` to guide your script-writing in this question.

- a) Connect to `bayes.biostat.washington.edu` using your favourite `ssh` client. Edit `submit_ex1_batch.sh` and `submit_ex2_batch.sh` to reduce the size of the simulations and confirm that you can get them to run as batch submissions.
- b) Conduct simulations using 3 additional values of n , of your choosing. Split each simulation into either 5 or 10 jobs (not 1 job and not 5000 individual jobs). Perform this batch submission either using a loop in your shell script, or by using a job array. **NB:** If `bayes` is full, `gosset` has an identical setup (i.e., use `qsub` to submit jobs) and is restricted to student use only. This is an older (i.e., slower) cluster, but often has available cores.

Present your results graphically and/or tabularly in a way that best illustrates your findings. Comment on what you see.

Upload your script(s) to your github repository along with your pdf write up of your results.