**Homework 4**

**Statistical Genomics**

**CROPS 545, Spring 2017**

Professor: Zhiwu Zhang

Due on March 22, 2017, Wednesday, 3:10PM, PST

**Objectives**: 1) Simulate phenotype from genotype; 2) GWAS by correlation; 3) Evaluate true and false positives.

**Comments:**

1. [ALPACA.R] It’s impressed to name the package delicately and meaningfully at the same time. Good job!
2. [ALPACA.R] It’s really great to prompt messages that show the progress.
3. [ALPACA.R, line 123] You can try to use data.frame() instead, which can construct the data list and name the variable in a single line. And data in the data frame is treated as variables, not columns. For example:

myGLM.results = data.frame (SNP = row.names(t(myGD))[-1],

Chromosome = myGM[,2],

Position = myGM[,3]…….

1. [ALPACA.R, line 127] It’s good to tell users how to inspect the results. But it would be better to print the list structure automatically, rather than ask users to use another command to print it out.
2. [ALPACA.R, line 127] The results won’t always be stored in the variable named “myGLM”, thus the message shown as “Use str(myGLM) to check the results” might confuse users. **(-1)**
3. Q3: Excellent user manual and tutorial.
4. Q3. Never demonstrate the demo code in the form of image, it’s difficult for user to follow your tutorial. **(-1)**
5. Q4: Good finding. But try to avoid printing signals in light yellow or light green, which are really hard to be distinguished from the background color.
6. Q5: It’s really great to show the difference by ROC curves in different scenario, excellent presentation.

Total point: 98