**Homework 4**

**Statistical Genomics (545)**

**Spring 2021**

Instructor: Zhiwu Zhang

Due on Friday, April 28, 2023, 3:10PM PST

**Hand in:** Email your R RMD and Knit PDF (or the link to the published on [preprint.org](https://www.preprints.org/) or [Peer Community Journal](https://peercommunityjournal.org/)) with subject of “StaGen545 HW4” to Zhiwu.Zhang@WSU.edu. Name your files as

“Homework4\_ firstname\_lastname”. The PDF file should have at least four pages. Your PDF file should focus on ideas, design, results, and interpretation by text, tables and figures. Leave the R code in the RMD file.

**Objectives**: 1) validation; 2) invalidate validation; 3) cross-validation; 4) GS, and 5) GWAS.

literature claims that incorporating GWAS results in genomic selection (GS) would improve accuracy than conducting GS and MAS along (e.g. two fold accuracy increase reported by Ravelombola et al., 2020). However, many of the reports have an invalidate procedure for such claim. As reviewed by McGowan et al. (2021) on Plant Breeding Reviewer (can be accessed by the links in Reference), invalidate procedure can artificially create fake accuracy improvement for such incorporation.

There are multiple ways to incorporate GWAS results in GS, including fitting the associated markers as covariates (Spindel , 2016) or building kinship using associated markers in gBLUP (Zhang et al. 2014). You can choose any format of incorporation to approve or disapprove the claim based on real phenotypes or simulated phenotypes. As MAS is in favor of Mendelian traits and GS polygene traits, your proof should be based on the comparisons of the incorporation to both GS and MAS (40 points).

Write a scientific paper including summary, introduction, results, method, discussion, conclusion, and references (60 points).

**Extra credit** (Due on May 7, 3pm): Submit data and source code to [GitHub](https://github.com/) (5 points), have manuscript on preprint such as [preprint.org](https://www.preprints.org/)(10 point) or publication on a journal including [Peer Community Journal](https://peercommunityjournal.org/) (20 points), and receive comments from reviewers (30 points).

**Reference**

1. Bernardo, R. 1994. **Prediction of maize single-cross performance using RFLPs and information from related hybrids**. Crop Sci. 34: 20–25.

[**https://doi.org/10.2135/cropsci1994.0011183X003400010003x**](https://doi.org/10.2135/cropsci1994.0011183X003400010003x)

1. Brian Rice and Alexander E. Lipka, **Evaluation of RR-BLUP Genomic Selection Models that Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum**, The Plant Genome, 2019, [doi: 10.3835/plantgenome2018.07.0052](https://doi.org/10.3835/plantgenome2018.07.0052).
2. Bertrand C.Y Collard and David J Mackill, **Marker-assisted selection: an approach for precision plant breeding in the twenty-first century**. Philosophical Transactions of the Rolyal Society B, 2007. <https://doi.org/10.1098/rstb.2007.2170>
3. Matthew McGowan, Jiabo Wang, Haixiao Dong, Xiaolei Liu, Yi Jia, Xianfeng Wang, Hiroyoshi Iwata, Yutao Li, Alexander E Lipka, and **Zhiwu Zhang**\*. **Ideas in Genomic Selection with the Potential to Transform Plant Molecular Breeding: A Review**. *Plant Breeding Review Vol 45* (eds. Irwin Goldman), John Wiley & Sons, Inc. 2021. pp. 273-320, [**https://doi.org/10.1002/9781119828235.ch7**](https://doi.org/10.1002/9781119828235.ch7), also available as [Preprint](https://www.preprints.org/manuscript/202010.0460/v2), [Publication](https://www.wiley.com/en-us/exportProduct/pdf/9781119828228), and [Google Book](https://www.google.com/books/edition/Plant_Breeding_Reviews_Volume_45/5j5IEAAAQBAJ?hl=en&gbpv=1&dq=Ideas+in+Genomic+Selection+with+the+Potential+to+Transform+Plant+Molecular+Breeding:+A+Review&pg=PA273&printsec=frontcover).
4. Spindel, JE, H Begum, D Akdemir, B Collard, E Redoña, J-L Jannink and S McCouch. **Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement**. Heredity,2016. <https://www.nature.com/articles/hdy2015113>.
5. VanRaden, P. M. 2008. **Efficient methods to compute genomic predictions**. J. Dairy Sci. 91: 4414–4423. [DOI: 10.3168/jds.2007-0980](https://doi.org/10.3168/jds.2007-0980)
6. Waltram Second Ravelombola, Jun Qin, Ainong ShiID, Liana Nice, Yong Bao, Aaron Lorenz, James H. Orf, Nevin D. Young, Senyu Chen, **Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation**. PLOS One, 2020, <https://doi.org/10.1371/journal.pone.0235089>.
7. Yao Zhou, MI Vales, Aoxue Wang, Zhiwu Zhang. **Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction**. *Briefings in Bioinformatics* 2016, <https://doi.org/10.1093/bib/bbx133>.
8. Zhe Zhang, Ulrike Ober, Malena Erbe, Hao Zhang, Ning Gao, Jinlong He, Jiaqi Li. **Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies.** PLOS One. <https://doi.org/10.1371/journal.pone.0093017>.
9. Zhang Z, Todhunter RJ, Buckler ES, Van Vleck LD. **Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood**. *J Anim Sci* 2007, **85**:881–885, <https://doi.org/10.2527/jas.2006-656>.