Statistical Genomics
CROPS 545
3 credit hours- Spring 2016

Professor: Zhiwu Zhang
Office 105 Johnson Hall.
Phone 509-335-2899
Email Zhiwu.Zhang@wsu.edu

Course Objective: To understand the statistical concept and principals in genomics for modern breeding programs.

Lecture: Johnson Hall 204. W/F 3:10-4:25 PM.

Office hours: M 1:30-2:30 PM

Attendance: Participation in each class is expected. Asking questions and contribution to discussions are part of lecturing theme in addition to homework and exams.

Course Description: Concepts and applications in modern breeding programs. The course includes three major sections: Fundamental, Genome Wide Association Study (GWAS) and Genomic Prediction/Selection (GS). The fundamental section covers the essential knowledge and skills of statistics, computer programming (R) and genomics. GWAS and GS sections cover the mechanisms, methods, and computing tools in GWAS and GS, respectively. We start from genotypes and pickup some of them as genes to simulate phenotypes. Then we examine how well we can map the genes and predict the phenotypes starting with very intuitive methods such as correlation and regression. Then we vary relevant factors to recognize their pitfall and strength. We also evolve statistical methods and computing tools all the way to their state of art, including mixed model and Bayesian methods. The course is beneficial for experimental design, data analyses to map genes controlling complex traits and predict their underlying genetic potential among individuals. Analytical skills, critical thinking and hand-on operations are emphasized throughout the teaching.

Required Text Book: There is no required textbook. Each lecture will be accompanied by a handout that covers all of the in class material and more in-depth material that is beyond this course. For students who would like to have a general reference book, I recommend:
Genome-Wide Association Studies and Genomic Prediction
http://link.springer.com/book/10.1007%2F978-1-62703-447-0

Prerequisites: Introductory statistics, linear algebra, computer programming, and genetics courses are recommended, but not required.

Grade:
Present (10%), Participant (10%), Midterm exam (20%), Final exam (30%) and Homework (30%). No late assignments will be accepted.

A (93%-100%); A- (90%-93%); B+ (87%-90%); B (83%-87%) B- (80%-83%); C+ (77%-80%); C (73%-77%); C- (70%-73%) D+ (66%-70%); D (60%-66%); F(0%-60%)

Exams

Midterm: February 26, Friday, 50 minutes (3:35-4:25PM), 25 questions.
Final: May 3, 120 minutes (3:10-5:10PM), 50 questions.

WSU Safety Information: alert.wsu.edu. Washington State University is committed to maintaining a safe environment for its faculty, staff, and students. Safety is the responsibility of every member of the campus community and individuals should know the appropriate actions to take when an emergency arises. In support of our commitment to the safety of the campus community the University has developed a Campus Safety Plan, http://safetyplan.wsu.edu. It is highly recommended that you visit this web site as well as the University emergency management web site at http://oem.wsu.edu to become familiar with the information provided.

Campus Resources

- Writing Center, http://www.writingprogram.wsu.edu/units/writingcenter/
- Library Services, http://www.wsulibs.wsu.edu/
- CACD, Center for Advising and Career Development, http://www.cacd.wsu.edu/
- Office of Student Conduct, http://conduct.wsu.edu
- Counseling and Testing Services, http://counsel.wsu.edu/
- Academic Integrity, http://academicintegrity.wsu.edu

Students with Disabilities: Reasonable accommodations are available for students with a documented disability. If you have a disability or may need accommodations to fully participate in this class, please visit the Access Center. All accommodations MUST be approved through the Access Center (Washington Building, Room 217). Please stop by or call 509-335-3417 to make an appointment with a disability specialist. http://accesscenter.wsu.edu

Plagiarism and Cheating policy (www.conduct.wsu.edu): Copyright and Intellectual Property WAC 504-25-018. **Academic dishonesty of any kind or degree will not be tolerated.** Academic integrity is a core component of graduate student and professional behavior. Any student caught cheating on any assignment will be given an F grade for the course and will be reported to the Office Student Standards and Accountability. Cheating is defined in the Standards for Student Conduct WAC 504-26-010 (3). It is suggested that you read and understand these definitions: http://app.leg.wa.gov/WAC/default.aspx?cite=504-26-010

Statistical Genomics

CROPS 545, Spring 2016

Lecture	Lecture	Section	Title	HW Due
1	1/13/16	Fundamental	Syllabus/course overview and introduction	
2	1/15/16		Computer programming in R	
3	1/20/16		Random variables and distribution	
4	1/22/16		Statistical inference	
5	1/27/16		Linear algebra[1]	
6	1/29/16		Genotyping By Sequencing (GBS)[2]	
7	2/3/16		Missing genotype imputation[3]	HW1
8	2/5/16		Genetic architecture and simulation of phenotype	
9	2/10/16		Linkage disequilibrium	
10	2/12/16	GWAS	GWAS by correlation	
11	2/17/16		Power, type I error and False Discovery Rate	HW2
12	2/19/16		Population structure and principal component analysis	
13	2/24/16		General Linear Model (GLM)	
14	2/26/16		Kinship	Midterm
15	3/2/16		Mixed Linear Model (MLM)[4]	HW3
16	3/4/16		Compressed MLM[6]	
17	3/9/16		Efficient Mixed Model Association (EMMA)[5]	
18	3/11/16		Population Parameter Previously Determined (P3D)[6,7]	
19	3/23/16		SUPER GWAS method[8,9]	HW4
20	3/25/16		Multiple Loci Mixed Model (MLMM)[10]	
21	3/30/16		FarmCPU	
22	4/1/16	Genomic Prediction	Marker Assisted Selection (MAS)	
23	4/6/16		Model fit and cross validation accuracy	
24	4/8/16		genomic Best Linear Unbiased Prediction (gBLUP)[11,12]	
25	4/13/16		Ridge regression (rrBLUP)	HW5
26	4/15/16		Kernel and machine learning	
27	4/20/16		Bayesian theory	
28	4/22/16		Bayesian methods[13]	
29	4/27/16		Bayesian implementation	
30	4/29/16		BLUP alphabet	HW6

Reference

1. Lynch, M. & Walsh, B. *Genetics and analysis of quantitative traits*. *Genetics and analysis of quantitative traits.* (1998).

2. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6,** e19379 (2011).

3. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11,** 499–511 (2010).

4. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38,** 203–208 (2006).

5. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178,** 1709–1723 (2008).

6. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42,** 355–360 (2010).

7. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42,** 348–354 (2010).

8. Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. A SUPER Powerful Method for Genome Wide Association Study. *PLoS One* **9,** e107684 (2014).

9. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* **8,** 833–835 (2011).

10. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44,** 825–830 (2012).

11. Zhang, Z., Todhunter, R. J., Buckler, E. S. & Van Vleck, L. D. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J. Anim. Sci.* **85,** 881–885 (2007).

12. VanRaden, P. M. Efficient methods to compute genomic predictions. *J Dairy Sci* **91,** 4414–4423 (2008).

13. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157,** 1819–1829 (2001).