Quantitative Genetics Graduate Education Module - Spring 2020

William A. Cresko

2020-04-07

2

Contents

Chapter 1

Course Overview

This is the complete set of *course materials* for the *Quantitative Genetics Graduate Education Module (GEM)* at the University of Oregon for the Spring of 2020. It is written in **Markdown** so that it can be easily updated.

In book you will find nearly all the information you will need to complete the course.

Chapter 2

Introduction to the course

This is the complete set of *course materials* for the *Quantitative Genetics Graduate Education Module (GEM)* at the University of Oregon for the Spring of 2020. It is written in **Markdown** so that it can be easily updated.

In this book you will find nearly all the information you will need to complete the course.

2.1 Instructor

Dr. Bill Cresko, wcresko@uoregon.edu

2.2 Course Information

Virtual Office Hours: M-W-F 11 to 12 (Zoom)

2.3 Software

- Latest version of R
- Latest version of RStudio

2.4 Inclusion and Accessibility

Please tell us your preferred pronouns and/or name, especially if it differs from the class roster. We take seriously our responsibility to create inclusive learning environments. Please notify us if there are aspects of the instruction or design of this course that result in barriers to your participation! You are also encouraged

to contact the Accessible Education Center in 164 Oregon Hall at 541-346-1155 or uoaec@uoregon.edu.

We are committed to making this course an inclusive and respectful learning space. Being respectful includes using preferred pronouns for your classmates. Your classmates come from a diverse set of backgrounds and experiences; please avoid assumptions or stereotypes, and aim for inclusivity. Let us know if there are classroom dynamics that impede your (or someone else's) full engagement.

Because of the COVID-19 pandemic, this course is being delivered entirely remotely. We realized that this situation makes it difficult for some students to interact with the material, for a variety of reasons. We are committed to flexibility during this stressful time and emphasize that we will work with students to overcome difficult barriers as they arise.

Please see this page for more information on campus resources, academic integrity, discrimination, and harassment (and reporting of it).

Chapter 3

Course Schedule

3.1 Background Material

- Introduction to quantitative genetics
- Reading materials
- Introduction to R and RMarkdown

3.2 Heritability

- Genetic variation
- Parent-offspring regression
- Line means analysis
- Environmental variation and trait distributions
- Covariation among traits
- The G-matrix

3.3 Selection Analysis

- Calculating fitness
- Selection analysis
- Correlation among traits
- Normalization of traits
- Calculation of selection gradients
- Comparison of the pattern of selection across populations

3.4 Quantitative Genetic Mapping

- Basic phenotyping and mapping information

- Quantitative Trait Loci (QTL) mapping
- Epistasis
- Genome Wide Association Studies (GWAS)

Chapter 4

Background material for the course

4.1 Introduction

Our goals in the first week are two-fold. The first is to become aquainted with R and RMarkdown so that you can begin to use it as an analysis platform for the rest of our work. The other goal is to become aquainted with the basic concepts of quantitative genetics. I'll start with links for the topical background and then provide links to help you get up to speed with R if you have not already been exposed to it. One of the secondary goals here is that you become comfortable enough with R to begin to use it in your own work.

4.2 Quantitative genetics

In class we will go over the basic elements of quantitative genetics in terms of relating variation at a single locus to the average atribute within a population and the variation among indviduals. We could obviously spend many months on this topic, but it will sufficient to read a couple of introductory papers on the subject.

If you ever feel the need to look into the population genetics background for our discussions, I highly recommend Graham Coop's online population genetics notes: http://cooplab.github.io/popgen-notes/

Here is a nice set of introdutory slides from Bruce Walsh that emphasize the points that I make in class: http://nitro.biosci.arizona.edu/workshops/GIGA/pdfs/Liege-2011-Intro-Quan-Gen.pdf

Graham also covers this information in narrative form, but also in a bit more

technical detail: http://cooplab.github.io/popgen-notes/#the-phenotypic-resemblance-between-relatives

Even a bit more technical are Bruce's lectre notes: http://nitro.biosci.arizona.edu/workshops/Uppsala2012/pdfs/Lecture03-Uppsala2012.pdf

If you want to dig in deeper, then Falconer and Mackay (1996) is a very readable introduction, although it is getting a bit out of date on the mapping side of things. (Good prices on used copies available.) https://www.amazon.com/Introduction-Quantitative-Genetics-Douglas-Falconer/dp/8131727408

If you want to go completely nuts then Lynch and Walsh (1998) and Walsh and Lynch (2018) deserve to be in your library, although they are boat anchors and quite expensive. (Good prices on used copies not generally available.) https://www.amazon.com/Genetics-Analysis-Quantitative-Traits-Michael/dp/0878934812/ref=pd_lpo_sbs_14_t_2?_encoding=UTF8&psc=1&refRID=G9RJZB2AMVME1R460BG3 https://www.amazon.com/Evolution-Selection-Quantitative-Traits-Bruce/dp/0198830874/ref=pd_lpo_sbs_14_t_1?_encoding=UTF8&psc=1&refRID=G9RJZB2AMVME1R460BG3

4.3   Matrix algebra

We will start talking about some multivariate topics next week, so it would be good if you brush up a bit on your basic matrix algebra. You may not have had to deal with these since pre-calculus. There are 100's of good resources for this on the internet. Here is one decent set of notes that are pretty compact. Mostly just concentrate on the basics of addition and multiplication for our purposes. You will need the more advanced topics as you move along in statistics and advanced topics in statistical genetics. http://www-bcf.usc.edu/~amarino/ML2.pdf

Here is a video of matrix multiplication if you like that sort of thing better. https://www.khanacademy.org/math/precalculus/precalc-matrices/multiplying-matrices-by-matrices/v/matrix-multiplication-intro

4.4   Introduction to R and RStudio

R is a core computational platform for statistical analysis. It was developed a number of years ago to create an open source envirnment for advanced computing in statistics and has since become the standard for statistical analysis in the field, replacing commerical packages like SAS and SPSS for the most part. Learning R is an essential part of becoming a scientist who is able to work at the cutting edge of statistical analysis – or to just do t-tests in a standard way. An important part of R is that it is script-based, which makes it easy to create reproducible analysis pipelines, which is an emerging feature of the open data/open analysis movement in science. This is becoming an important com-

ponent of publication and sharing of research results, so being able to engage fully with this effort is something that all young scientists should do.

RMarkdown is an extra layer placed on top of R that makes it easy to integrate text explanations of what is going on, native R code/scripts, and R output all in one document. The final result can be put into a variety of forms, including webpages, pdf documents, Word documents, etc. Entire books are now written in RMarkdown and its relatives. It is a great way to make quick webpages, like this document, for instance. It is very easy to use and will be the format that I use to distribute your assignments to you and that you will use to turn in your assignments.

It sounds like most of you have had some exposure to R. All of you need to download R onto your computers for use. Get that here: https://www.r-project.org/

I also strongly recommend that you download RStudio, as it will make running and documenting your assignments much easier. It is a great frontend for building RMarkdown documents as well Get that here: https://www.rstudio.com/

4.4.1 Learning resources

There are tons of resources for learning R and RMarkdown on the internet. I won't try to capture them here.

There is an organized group that is dedicated to training in R called DataCamp (https://www.datacamp.com/). They provide all of the basics for free. They actually have training for most data science platforms. RStudio provides links for training directly related to R and RMarkdown here: https://www.rstudio.com/online-learning/

There are also many, many R training videos on YouTube. Most of them are very well meaning but fairly amateurish.

You can also go the old "paper" manual route by reading the materials provided by R itself: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

In reality, if you want to do almost anything in R, simply type in what you are interested in doing into Google and include "in R" and a whole bunch of links telling you exactly what to do will magically appear. Most of them appear as discussions on the StackOverflow website. In that case, the first thing that you see is the question–usually someone doing it just a bit wrong–so you should scroll down to see the right way to do it in the answers. It is really an amazing resource that will speed you along in nearly every form of analysis that you are interested in.

Please do not hesitate to contact me if you have any questions or run into any obstacles. The point of this class is to learn by doing, but my aim is that the doing should not be too hard at this point.

Chapter 5

Phenotypic traits and values

5.1 Introduction

Many phenotypes of organisms are complex, meaning that they are the product of multiple loci interacting with one another and the environment to create smooth distributions.

Chapter 6

Background material for the course

6.1   Introduction

Our goals in the first week are two-fold. The first is to become aquainted with R and RMarkdown so that you can begin to use it as an analysis platform for the rest of our work. The other goal is to become aquainted with the basic concepts of quantitative genetics. I'll start with links for the topical background and then provide links to help you get up to speed with R if you have not already been exposed to it. One of the secondary goals here is that you become comfortable enough with R to begin to use it in your own work.

6.2   Quantitative genetics

In class we will go over the basic elements of quantitative genetics in terms of relating variation at a single locus to the average atribute within a population and the variation among indviduals. We could obviously spend many months on this topic, but it will sufficient to read a couple of introductory papers on the subject.

If you ever feel the need to look into the population genetics background for our discussions, I highly recommend Graham Coop's online population genetics notes: http://cooplab.github.io/popgen-notes/

Here is a nice set of introdutory slides from Bruce Walsh that emphasize the points that I make in class: http://nitro.biosci.arizona.edu/workshops/GIGA/pdfs/Liege-2011-Intro-Quan-Gen.pdf

Graham also covers this information in narrative form, but also in a bit more

technical   detail: http://cooplab.github.io/popgen-notes/#the-phenotypic-resemblance-between-relatives

Even a bit more technical are Bruce's lectre notes: http://nitro.biosci.arizona.edu/workshops/Uppsala2012/pdfs/Lecture03-Uppsala2012.pdf

If you want to dig in deeper, then Falconer and Mackay (1996) is a very readable introduction, although it is getting a bit out of date on the mapping side of things. (Good prices on used copies available.) https://www.amazon.com/Introduction-Quantitative-Genetics-Douglas-Falconer/dp/8131727408

If you want to go completely nuts then Lynch and Walsh (1998) and Walsh and Lynch (2018) deserve to be in your library, although they are boat anchors and quite expensive. (Good prices on used copies not generally available.) https://www.amazon.com/Genetics-Analysis-Quantitative-Traits-Michael/dp/0878934812/ref=pd_lpo_sbs_14_t_2?_encoding=UTF8&psc=1&refRID=G9RJZB2AMVME1R460BG3 https://www.amazon.com/Evolution-Selection-Quantitative-Traits-Bruce/dp/0198830874/ref=pd_lpo_sbs_14_t_1?_encoding=UTF8&psc=1&refRID=G9RJZB2AMVME1R460BG3

6.3   Matrix algebra

We will start talking about some multivariate topics next week, so it would be good if you brush up a bit on your basic matrix algebra. You may not have had to deal with these since pre-calculus. There are 100's of good resources for this on the internet. Here is one decent set of notes that are pretty compact. Mostly just concentrate on the basics of addition and multiplication for our purposes. You will need the more advanced topics as you move along in statistics and advanced topics in statistical genetics. http://www-bcf.usc.edu/~amarino/ML2.pdf

Here is a video of matrix multiplication if you like that sort of thing better.   https://www.khanacademy.org/math/precalculus/precalc-matrices/multiplying-matrices-by-matrices/v/matrix-multiplication-intro

6.4   Introduction to R and RStudio

R is a core computational platform for statistical analysis. It was developed a number of years ago to create an open source envirnment for advanced computing in statistics and has since become the standard for statistical analysis in the field, replacing commerical packages like SAS and SPSS for the most part. Learning R is an essential part of becoming a scientist who is able to work at the cutting edge of statistical analysis – or to just do t-tests in a standard way. An important part of R is that it is script-based, which makes it easy to create reproducible analysis pipelines, which is an emerging feature of the open data/open analysis movement in science. This is becoming an important com-

ponent of publication and sharing of research results, so being able to engage fully with this effort is something that all young scientists should do.

RMarkdown is an extra layer placed on top of R that makes it easy to integrate text explanations of what is going on, native R code/scripts, and R output all in one document. The final result can be put into a variety of forms, including webpages, pdf documents, Word documents, etc. Entire books are now written in RMarkdown and its relatives. It is a great way to make quick webpages, like this document, for instance. It is very easy to use and will be the format that I use to distribute your assignments to you and that you will use to turn in your assignments.

It sounds like most of you have had some exposure to R. All of you need to download R onto your computers for use. Get that here: https://www.r-project. org/

I also strongly recommend that you download RStudio, as it will make running and documenting your assignments much easier. It is a great frontend for building RMarkdown documents as well Get that here: https://www.rstudio.com/

6.4.1 Learning resources

There are tons of resources for learning R and RMarkdown on the internet. I won't try to capture them here.

There is an organized group that is dedicated to training in R called DataCamp (https://www.datacamp.com/). They provide all of the basics for free. They actually have training for most data science platforms. RStudio provides links for training directly related to R and RMarkdown here: https://www.rstudio. com/online-learning/

There are also many, many R training videos on YouTube. Most of them are very well meaning but fairly amateurish.

You can also go the old "paper" manual route by reading the materials provided by R itself: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

In reality, if you want to do almost anything in R, simply type in what you are interested in doing into Google and include "in R" and a whole bunch of links telling you exactly what to do will magically appear. Most of them appear as discussions on the StackOverflow website. In that case, the first thing that you see is the question–usually someone doing it just a bit wrong–so you should scroll down to see the right way to do it in the answers. It is really an amazing resource that will speed you along in nearly every form of analysis that you are interested in.

Please do not hesitate to contact me if you have any questions or run into any obstacles. The point of this class is to learn by doing, but my aim is that the doing should not be too hard at this point.