Syllabus

AGRON 5130 - Data Science for Agricultural Professionals

Instructor: Dr. Caio dos Santos

E-mail: clsantos@iastate.edu

Office Hours: Office hours will be held by appointment. Please e-mail me to schedule office hours. Key activities include answering unclear content from modules, review of key module content, and assignment issues/questions.

Course Overview

This course is strongly influenced by the former instructor's industry experience. Dr. Harbur taught this class between 2012 and 2025, while also working as a full-time statistician in the agricultural industry.

We will start with how to summarize yield results from a yield map, then work our way through sampling from multiple fields, "side-by-side" trials, and more extensive hybrid or product comparisons. We will also look at how quantitative treatments (like different product rates) can be modeled, and yield responses predicted. We will learn how to present data, and how spatial data is used to create fertilizer application maps. And we will spend a unit on machine learning, to better understand how "Ag Tech" tools you use make predictions and find patterns in data.

In this course, we will use the **R** statistical language, which is an open-source programming language. This means that the skills you develop here can be used beyond this class, as you will not need a license to analyze data.

Course Description

Quantitative methods for analyzing and interpreting agronomic information. Principles of experimental design, hypothesis testing, analysis of variance, regression, correlation, and graphical representation of data. Use of R and R Studio for organizing, analyzing, and presenting data. Required course for the Master of Science in Agronomy degree program.

Course Prerequisites:

AGRON 1810 or equivalent, MATH 1400, STAT 1040. Restricted to graduate students enrolled in MS Agronomy online degree program at ISU. Students from other departments must get permission.

Look for my welcome message in Canvas to learn more about my background. You will introduce yourself as part of Discussion Topic 1.1.

Course Philosophy

The purpose of this course is to learn how to apply various statistical tools to design experiments and interpret data. We want to focus on understanding and using the tools, rather than only memorizing formulas and definitions. Therefore, the best strategy for success may vary somewhat for this course from those used for other courses.

A useful analogy is learning to play a musical instrument. You would not expect to be able to learn how to play the piano by reading about it in a book. Of course, you could learn the theory behind the instrument, but translating that into the ability to play would be a stretch. To learn to play it, you have to ... practice. To be an accomplished pianist you have to practice a lot.

Course Objectives (COs)

1. Calculate, use, and interpret models to describe the distribution of population and sample data.
2. Formulate and test hypotheses about the differences among two or more populations, using t-tests, analyses of variance, and post-hoc means-separation tests.
3. Identify optimal experimental designs to minimize experimental error or accommodate plot treatment logistics and apply data-cleaning techniques to messy data.
4. Estimate the association or causation between continuous variables using simple linear, multiple, and non-linear regression models.
5. Explain how spatial statistics and machine learning can be used to address more complex modeling needs of agricultural data.

Course Structure

The online course materials in Agronomy 5130 consist of 14 modules. There are no exams in this class.

Each module is designed to take one week to complete. Your weekly activities will typically include the following:

1. Reading the online modules and utilizing the included learning tools.
2. Reading any required textbook pages.
3. Completing and submitting each homework assignment. Homework assignments allow you to practice the concepts you have learned in each module. I will allow one re-submission of every homework in which you can incorporate my suggestions on how to better solve problems. If you get 100% of your first try, you will not need this.
4. Responding thoughtfully to each Discussion Topic. There are ten discussion topics that are assigned throughout the semester. You will be randomly divided into subgroups of 4 – 6 students for each topic. These questions are designed to explore important principles. Not every module has a discussion topic, and Module 8 has two discussions. To get full credit for each discussion topic, post your own answer to the discussion question and write one response to another group member's post.
5. Completing and submitting each Module Reflection. At the end of each module, you will be asked to respond to three questions:

- In your own words, write a short nontechnical summary (about 150 words) for this module. Please be specific about the module content. The summary need not be perfect, it is an exercise to help you synthesize the material.
- List the most valuable concept you've learned from the module. This helps us to identify what areas to expand or reduce in future versions of the course.
- Describe module concepts that are hardest to understand. This helps us understand what sections may need more work.

Required Textbook

The textbook used in this class was written for this course by the former instructor. The book is hosted online and can be found in the link below. In addition, a copy of the book has also been added to the Canvas course page.

Harbur, Marin 2021. Data Science for Agricultural Professionals. Open source e-book, available online at: https://incomparable-valkyrie-81dd57.netlify.app/

Required Technology

We will use the programming language **R** (https://cran.rstudio.com/) for this course. There are many ways to interact with this programming language (e.g., graphical user interfaces, text editors) but for this class I will ask you to use **RStudio** (https://posit.co/download/rstudio-desktop/), which contains a nice user interface that works really well with **R**. Please check the course introduction section on Canvas, in which I will include a video of how to install this software.

Assignments and Grading

Grading Procedures

Letter grades for the course will be assigned as follows:

A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	74-76
C-	70-73
D+	68-69
D	65-67
F	<65

Expectations

Homework

The homework assignments in this course are designed to let you practice the skills and techniques that have been presented to you. The best way to master the course material and learn the methods is to do the

homework. Any shortcuts you take there will show up later when you are asked to apply the methods on your own. Please do not shortchange yourself by turning in the work of someone else as your homework. This hampers your learning, is unethical and represents academic dishonesty. Homework is part of the learning process. **Homework will be due on Monday of the following week at 8 AM (US Central Time).**

Module Reflections

At the end of each module, you will be asked to respond to three questions:

1. In your own words, write a short summary (150 words) of each week's module. Be specific about the module content, not your general feelings about the module (save that for the third question), but represent the entire module. I want to be sure that you have kept up with the material.
2. What is the most valuable concept you learned from the module? Why is this concept valuable to you?
3. What concepts in the module are still unclear/the least clear to you? These help me develop the Questions and Answers section at the end of each module!

Module reflections are an essential way for us to communicate. I read your summaries to assess whether or not you understand the material and will let you know if you make a mistake.

Please answer Question 3 each week. I know your answers to Question 3 may be very similar from week to week, but please do not answer "the same as last time".

Discussions

There are ten discussion topics that are assigned throughout the semester. I realize that, in many cases, there is less to discuss after the first few students have made a post. However, the questions are designed to make you think about important principles and I would like for you to be as thoughtful and creative as possible in answering them. You may be surprised what you can learn from each other by taking the time to view posts and respond to them. To get full credit for each discussion topic, you must post to the topic, respond to at least one post from another student, and read at least 10 responses made by other students.

Quizzes

Each quiz will cover the previous three modules. There are four quizzes in this course, covering Modules 1 - 3, 4 - 6, 7 - 9, and 10 - 12. They are open-book and untimed, and you will have over a week to complete them. While memorization is great, it is more important you learn how to find your way back to answers. I hope this approach will reduce your stress.

Policies

Communication Policy

All communication within the course should adhere to University standards of Netiquette at ISU. Specifically, communication should be scholarly, respectful, professional, and polite. You are encouraged to disagree with other students, but such disagreements need to be based upon facts and documentation. It is my goal to promote an atmosphere of mutual respect in our interactions. Please contact me if you have suggestions for improving the interactions in this course.

Use the Course Questions forum in Canvas (or the Canvas Inbox) to ask questions, share an interesting article or observation, or comment on current and relevant events. Keep informed—check the discussion

board frequently. I will monitor the discussion board and my Canvas Inbox during "regular business hours" and you can expect a timely response (usually within one to two business days). I will not be responding to email after Friday at 5pm until Monday morning.

General announcements will be posted to the Announcements section of Canvas. Additional guidelines apply to communication within Discussion Topics. Please review the Discussions shown above.

Be sure to properly configure your Notification settings or commit yourself to checking Canvas daily for new communication.

Feedback Policy

All graded assessments will be returned with feedback usually within 7-14 days of the due date. Be sure to check your graded submissions in Canvas for comments regarding your work.

Deadlines

All deadlines are posted on the Course Calendar in Canvas. Need extra time to meet a deadline? Explain the situation to your instructor **in advance** and accommodations can be made. The explanation doesn't need to be extensive. The important factor is to attempt to notify the instructor ahead of time.

Tentative Schedule

Week	**Content**
Week 1: Jan 20 - 23	Introduction to R programming
Week 2: Jan 26 - 30	Population Statistics
Week 3: Feb 2 - 6	Distributions and Probability
Week 4: Feb 9 - 13	Sample Statistics
Week 5: Feb 16 - 20	Two-Treatment Comparisons
Week 6: Feb 23 - 27	Hypotheses Testing
Week 7: Mar 2 - 6	Multiple Treatment Trials
Week 8: Mar 9 - 13	Multiple Treatment Designs
Week 9: Mar 16 - 20	Means Separation and Data Presentation
Week 10: Mar 23 - 27	**Spring Break**
Week 11: Mar 30 - Apr 03	Messy and Missing Data
Week 12: Apr 6 - 10	Correlation and Simple Regression
Week 13: Apr 13 - 17	Nonlinear Relationships and Multiple Linear Regression

Week	Content
Week 14: Apr 20 - 24	Spatial Statistics
Week 15: Apr 27 - May 1	Machine Learning
Week 16: May 4 - May 8	Putting it All Together