Statistics 5200 Analysis of Designed Experiments Fall 2020

Key Points of Syllabus

- This course emphasizes statistical thinking over hand calculation or algebraic manipulation
- Notes posted (and software links) on class website (link from Canvas)
- Homework posted on and submitted through Canvas (only PDFs) – please respect (but don't fear) due dates
- Class meeting behavior will exhibit mutual respect – feel free to ask questions and participate
- Grades not inflated or curved, and based on homework and exams
- Grading only discussed in office hours or via email
- We will use SAS; help available in office hours (or by appointment) and (to some degree) via email
- Students are trusted to read the rest of the syllabus

Class Meetings: MWF 10:30-11:20 via Zoom – see Canvas for link
Course Website: www.stat.usu.edu/jrstevens/stat5200 (link from Canvas)

Instructor: Dr. John R. Stevens
Contact: 797-2818, `john.r.stevens@usu.edu`
Office Hours: MW 11:30am - 1:00pm (note: not Fridays) via Zoom (link on Canvas), or by appointment

Prerequisites: An introductory statistics class, such as STAT 2000 / 3000 or WILD 6500.

Textbook: A First Course in Design and Analysis of Experiments; freely available online at `http://users.stat.umn.edu/~gary/Book.html` (link on course website).

Course Description: This class is an introduction to the design and analysis of balanced experiments. In this class the emphasis will be placed on the analysis of real data sets from many disciplines using current computer software. How to choose the appropriate statistical models to analyze a given data set, and how to interpret the results of a statistical analysis in the context in which the experiment was conducted, will be emphasized over hand calculation and algebraic manipulation. By the end of this class you will be able to recognize most of the commonly used experimental designs, including factorial designs, block designs, split-plot and split block designs, repeated measures designs, and nested designs. You will be able to identify situations in which each of these kinds of designs is applicable. You will be able to construct and analyze each of these kinds of design.

Student Responsibilities: Attend class, participate in class discussions (off-task behavior is not acceptable), and complete assignments on time. Be responsible for all material presented in class as well as material covered by the assigned homework problems. Feel free to ask questions and give answers in a positive, non-threatening classroom environment. Take advantage of office hours to receive additional instruction as necessary.

Office Hours: While questions and discussions are welcome and expected during class periods, all students should feel free to visit office hours for individual assistance with the course material.

Students unable to attend office hours may make an appointment to see the instructor at another time. Students may not access their email on the instructor's office computer.

Grading: Your grade will be based on homework assignments (40%), a midterm exam (20%), and a final exam (40%). Anticipated grade cut-offs are 94 A, 90 A-, 87 B+, 83 B, 80 B-, 77 C+, 73 C, 70 C-, 65 D+, 60 D. If every student does well, then every student can get an A. Scores and grades will not be inflated and are only an indication of what you do, not who you are. Questions regarding grades or scores will **only** be considered by email (not Canvas) or during office hours.

Homework: Some assignments may be worth more than others. Submit completed assignment as a single document through Canvas (http://canvas.usu.edu) before 8 P.M. on the due date. Late work will be accepted, but may not be graded as promptly. (Please don't go far past the due date unless you have extenuating circumstances, the due dates are for your own good and the grader's sanity.) Only PDF file submissions will be allowed. Homework assignments should be typed neatly with necessary computer output (but not code) placed in order with each corresponding homework exercise. Figures (including fonts) should be clear and readable. Each homework should include an appendix with relevant SAS code. Any unnecessary computer output (which is common with SAS) will result in points deducted. Key numerical results in pasted output should be highlighted, and for many questions, short bullet-point answers (as opposed to long ramblings) will be perfectly acceptable. Homework handed in must be the student's own work.

Starting Homework: For each Homework assignment, think first what needs to be done (conceptually and then statistically), and then construct (maybe using available sample code) the appropriate SAS code to perform the analysis. If you just blindly modify available sample SAS code, you will waste a lot of time (and not learn as much). This will require you to first understand what the available sample SAS code does, so you can find which part(s) of it you will need to modify and use (or not).

Exams: Take-home midterm and final exams will be given. Dates and details will be announced in class and on Canvas.

Handouts and Course Notes: Print off or save handouts from the course website. Even though some handouts have color, they can be printed in black and white without loss of information.

Note: Students with ADA-documented physical, sensory, emotional or medical impairments may be eligible for reasonable accommodations. Veterans may also be eligible for services. All accommodations are coordinated through the Disability Resource Center (DRC) in Room 101 of the University Inn, (435)797-2444. Please contact the DRC as early in the semester as possible. Alternate format materials (Braille, large print, digital, or audio) are available with advance notice.

SAS: The SAS computer package is available in several campus computer labs, and you will also have access to a free online version during the semester. There are many software packages that could be used, and the textbook refers to several, but few (if any) are as comprehensive and widely used as SAS, particularly for experimental design. SAS is so widely used that good programming skills in SAS are sufficient to get one a good job in industry. Because it is so widely used and because it is such a good statistical package, SAS will be used in this course.

General Course Background: The origins of experimental design lie in agricultural experimentation. Many of the designs we will discuss were discovered (or formalized) by two famous statisticians, R.A. Fisher and F. Yates, at Rothampstead Agricultural Station in England in the 1920's through 1940's. These designs (and their underlying principles) are widely used in agriculture, all the life and medical sciences, all the social sciences, business, industry, and in manufacturing everything from snack foods to jet engine components. The applications of these principles in industrial processes have been an important aspect of the quality improvement revolution in many countries.

Topics and Tentative Schedule: The following list summarizes main topics to be covered in the course, along with an approximate amount of time for each:

- **A: Experiments and Terminology of Design**
  (2 weeks; Chapters 1-2 in the textbook.)
  What do we mean by an experiment? Why do an experiment? Experimental units, treatments, factors, and other terms. Randomizing to eliminate bias and confounding; randomizing run-orders and other things; randomization as a basis for statistical inference; paired and 2-sample t-tests.

- **B: Completely Randomized Design**
  (2.5 weeks; Chapters 3-6 in the textbook and other sources.)
  Random variables; probability distributions; means, variances, and covariances; models; summary statistics; transforming to normality. The fixed effects model; the ANOVA decomposition and why ANOVA works; post hoc mean comparisons; contrasts; residual analysis for validating assumptions.

- **C: Factorial Designs**
  (3 weeks; Chapters 8-9 in the textbook.)
  The analysis of experiments with two or more fixed treatment factors. Fixed effects models; main effects versus interactions; characterizing interactions; interpretation of F-ratios; contrasts and post hoc mean comparisons; unreplicated designs.

- **D: Random, Mixed, and Nested Effects**
  (2.5 weeks; Chapters 11-12 in the textbook.)
  Random effects models; ANOVA for random effects; variance components; expected mean square algorithm; exact and approximate F-tests for random factors; Mixed effects models; the expected mean square algorithm; exact and approximate F-tests; REML estimation and the MIXED procedure in SAS. Nested and nested factorial designs; expected mean squares; estimation of variance components by REML.

- **E: Designs with Restrictions on Randomization**
  (Approximately 2.5 weeks; Chapters 13 and 16 in the textbook and other sources.)
  Blocking; the randomized complete block design; Latin squares and Graeco-Latin squares; cross-over designs. Whole plot, sub-plot, and sub-sub-plot structure and restricted randomization; split-plot, split-split plot, and split-block designs; repeated measures designs; covariance structures for repeated measures designs; crossover designs.

- **F: Additional Topics**
  (2 weeks; Certain skipped chapters in the textbook and other sources.)
  Power and sample size calculations, generalized linear mixed models, ANCOVA, incomplete and fractional designs, restricted vs. unrestricted mixed models.