Syllabus for STAT 3301: Statistical Modeling for Discovery I
Autumn 2024 – 3 credit hours

Instructor: Dr. Andrew Richards
BS, Physics, U.S. Naval Academy
MBA, Yale School of Management
PhD, Statistics, The Ohio State University
Office: Cockins Hall 325
Office Hours: MWF 1:15-2:15 and by appointment
E-mail: richards.1227@osu.edu **not buckeyemail**
TA: Sooyon Kim, Kim.9462@buckeyemail.osu.edu.

Course meeting times and locations: MWF 10:20-11:15pm or 11:30-12:25 in Pomerene 150

Prereq: C- or above in 3202; or 4202 and 5730; or permission of instructor. Prereq or concur: Math 2568, or permission of instructor.

Required Text: *Applied Linear Regression*, Fourth Edition (2014) by Sanford Weisberg.

An electronic version of the text can be accessed for free through The Ohio State University Libraries at `https://library.ohio-state.edu/record=b7651844~S7`. You will need to click on "Connect to resource EBSCOhost"; you may also need to supply your OSU credentials. The online resource is best suited for screen reading; each individual is allowed to print/e-mail/save/download a limited number of pages.

Required software:

- This class requires you to use the free statistical software package called R (The R Project for Statistical Computing; `http://www.r-project.org/`).
  - You can download R for Windows, Mac, and Linux, from the CRAN archive at `https://cran.r-project.org`.
  - An in-depth introduction to R is available at `http://cran.r-project.org/doc/manuals/Rintro.pdf`
  - Hands-on tutorials are available in the Swirl system, which you can learn about at `http://swirlstats.com/`. In particular, "R Programming: The basics of programming in R" is an appropriate first tutorial for students who have never used R.
- An easier to use interface to R is available in the software package RStudio. This package is available for Windows, Mac, and Linux and can be downloaded for free from `http://rstudio.org`. Note that RStudio requires R to be installed.
- This class requires the use of the (free) R Markdown authoring framework to complete assignments. Information about R Markdown will be provided in class; an online guide with overview information can be found at `https://rmarkdown.rstudio.com`.

Website: Please visit `http://www.carmen.osu.edu/`. Carmen is used extensively for this course, so you should check daily for announcements about the class and other class material. Contact the IT Service Desk at 614-688-4357 (HELP) for help with access.

Course Description: Statistical models for data analysis in the linear regression framework. The challenges of developing meaningful models for data are explored, with emphasis on the model building process, the use of numerical and graphical diagnostics for assessing model fit, and interpretation and communication of results. Statistical foundations are introduced along with basic inferential techniques.

Learning Outcomes: By the end of this course, students should successfully be able to:

- Use graphical and numerical summaries of data to describe relationships between variables.
- Formulate, fit, evaluate, and compare regression models that describe relationships between variables.
- Understand and be able to describe the statistical foundations of standard regression models.
- Identify common violations of the assumptions that underly standard regression models.
- Perform a complete regression analysis and communicate the results in both statistical and problem-specific terms.
- Distinguish between descriptive and causal interpretations of regression.

Homework:

Description: There will be ten homework assignments. Homework problems that require R software should be completed in R Markdown and a knitted html file should be uploaded. Homework problems that do not require R may be handwritten (electronically, or on paper and scanned) and uploaded to Carmen by the due date. All work and software output should be uploaded as a single pdf file unless stated otherwise.

Academic integrity and collaboration: The purpose of the written homework is to assess and provide feedback on your understanding. **Therefore, answers with little or no explanation or work shown will receive no credit.** For the homeworks as well the exams, your solution should be clear and detailed to explain your understanding of the course.

While grading the homeworks, it may not be possible for us to provide detailed explanations on each question that is graded. To make up for this, I will endeavor to create homework solutions that are detailed enough to allow you to understand how the question could be approached. You may consult with other students, however, the **work submitted must be your own.**

Data analysis assignment:

Description: here will be an individual, comprehensive data analysis project that will be completed by the end of the semester. Expect details on the project to be posted in early November.

Academic integrity and collaboration: The data analysis project is individual and should be treated as such. It should be completed **without any external help or communication.** Sharing of code or other discussion between students is **strictly prohibited**.

Exams:

Description: There will be three midterm exams. The midterms will be held during lecture on the dates listed in the schedule.

Academic integrity and collaboration: You must complete the midterm and final exams yourself, **without any external help or communication.** Sharing of any items such as calculators or formula sheets is **prohibited**. Again, answers with little or no explanation or work shown **will receive no credit**. Students are **strongly advised** to prep a formula highlight sheet in advance.

Late assignments policy:
Assignment solutions will be posted shortly after the submission deadline. No late assignments will be accepted without **prior permission** and/or **formal documentation**. Please refer to Carmen for due dates. Accommodations can be made in case of severe illness, so please notify me as soon as possible if this situation arises. Deadlines are crucial in order, among other things, to:

- Get grading done and provide feedback in a timely manner
- Grade all assignments at the same time to maintain consistency and fairness
- Provide a mechanism to help ensure students keep up with the material and are prepared for follow-on lectures
- Protect students from their inability to predict their own future behavior – "I'll somehow manage to catch up at the end of the semester."

Course attendance policy: You are expected to attend all lectures. I will take attendance at lecture, and students are responsible for all material covered in class. I intend to simulcast lectures on zoom for students who are sick, but **I will not record lectures or provide annotated notes**. Students should keep all electronics closed during class with the exception of taking notes on a tablet. Office hours should not be used for instruction on material that has already been covered in class.

Course technology: In addition to R software, students are expected to have a basic working knowledge of The Microsoft Office software. All Ohio State students are now eligible for free Microsoft Office 365. Visit the `go.osu.edu/office365help` help article for full instructions.

Final Grade: Your final course grade will be based on the following weighting of assessment components:

Category	**Percentage**
Homework	20
Exam 1	20
Exam 2	20
Exam 3	20
Final project	30
Less weakest category	-10
Total	100

Grading Scale:
Grades will be assigned according to the scale below, with course components weighted as listed above.

93-100 = A
90-92.9999 = A-
87-89.9999 = B+
83-86.9999 = B
80-82.9999 = B-
77-79.9999 = C+
73-76.9999 = C
70-72.9999 = C-
67-69.9999 = D+
60-66.9999 = D
< 60 = E

E-mail Correspondence: In order to protect your privacy, all course email correspondence must be done through a valid OSU name.nn account. Please use the correct email address. (Richards.1227@osu.edu **not** @buckeyemail.osu.edu). Please write "STAT 3301" somewhere in the subject line, as this will help me to quickly identify and reply to class emails. It is reasonable to expect a response within one business day.

Other information: Other standard university boilerplate can be found here: `https://asccas.osu.edu/curriculum/syllabus-elements`.

Copyright: The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Disclaimer:
The planned instruction for this course may be disrupted for a number of reasons. Such disruptions may affect individual students for a brief period of time, the entire class, the instructor, or the entire university. If the class is disrupted, we will adjust as needed. The adjustments may include changes to course delivery, assignments, grading of assignments, and determination of final course grade. Please pay special attention to announcements in class and over Carmen. **Failure to address every possible scenario in this syllabus does not override your responsibility to exercise basic common sense. If in doubt about any course policy, ask in advance!**

Acknowledgemnt:
Thank you to Dr. Steephanson Anthonymuthu for his kind sharing of advice and course materials in preparation for this semester.

Tentative Course Schedule

Week	Dates	Topics, Holiday, Homework, and Exam Dates
1	8/27-8/29	Intro.
2	9/1-9/5	Correlation. **Labor Day 9/1**
3	9/8-9/12	SLR model. **HW1 9/8**
4	9/15-9/19	Inference and prediction with SLR. **HW2 9/15**
5	9/22-9/26	Sums of squares **HW3 9/24 MT1 9/26**
6	9/29-10/3	Diagnostics and transformations. **HW4 10/3**
7	10/6-10/10	MLR model. **HW5 10/10**
8	10/13-10/17	Polynomial regression. **Fall Break 10/17**
9	10/20-10/24	MLR inference. **HW6 10/20**
10	10/27-10/31	Confidence and prediction intervals. **HW7 10/27 MT2 10/29**
11	11/3-11/7	Categorical predictors. **HW8 11/7**
12	11/10-11/14	Categorical and continuous predictors. **Veterans Day 11/11**
13	11/17-11/21	Multifactor models. **HW9 11/17**
14	11/24-11/28	Model comparison. **Thanksgiving 11/26-11/28**
15	12/1-12/5	Cross validation. **HW10 12/1 MT3 12/3**
16	12/8-12/10	Stepwise and nonlinear regression.

FINAL PROJECT DUE: 11:59pm, Tuesday, December 16.