MSAN 601 – Linear Regression Analysis
Instructor: Paul Intrevado
Course Syllabus
Fall 2016

SUMMARY INFORMATION
Office: SFH 525, 101 Howard Campus
Office Hours: open door policy, by appointment, videoconference or Slack
Mobile Phone: 765.418.6874
Office Phone: 415.422.2527
Email: pintrevado@usfca.edu

Class Location: SFH 529, 101 Howard Street
Class Time: 09:30 - 11.30 / 13.30 - 15.30 Mondays and Wednesdays
TA Hours: 10:00 - 12.30 Fridays in the 5^{th} Floor Agora at 101 Howard

ON COURSE GOALS. Any student who successfully completes this course should:

- Understand the structural forms of classical simple and multiple linear regression models, as well as the assumptions underlying these models;
- Be familiar with the principles underlying parameter estimation for these classes of models;
- Formulate and test hypotheses, as well as use models for both prediction and explanation;
- Be able to use `R` to load and manipulate data, fit regression models, and generate various outputs like ANOVA tables, confidence intervals for parameters, diagnostic assessments, etc.;
- Be able to verify/test whether or not fitted residuals conform to the assumptions that underlie classical regression;
- Rigorously identify and manage outliers and influential observations, all while understanding the potential implications of leaving them uncensored;
- Be able to modify regression models to encourage fitted residuals to more closely conform to the assumptions that underlying classical regression, i.e., handle multicollinearity, heteroscedasticity, autocorrelation, a non-normal error term, and specification errors;
- Understand how to use dichotomous dependent variables, indicator (or dummy) independent variables, and interaction terms;
- Place regression analysis within a business context, i.e., as an analyst who dexterously leverages regression as a tool rather than an analyst who treats regression as a black box; and
- Have a sufficiently intimate knowledge of the mathematical mechanics underpinning linear regression modeling so as to be able to commit any component of linear regression analysis to `R` code, as opposed to using an existing `R` package
- Be able to communicate the results of complete and well-reasoned regression analysis both orally and in writing.

ABOUT ME. My name is Paul Intrevado. Please call me Paul. I'm an Assistant Professor of Analytics in the Department of Mathematics & Statistics, in the College of Arts & Sciences. I am also the Director of the Practicum Program for MSAN as well as the Associate Director of the Data Institute. Please feel free to knock on my office door or call/text/Slack me at your convenience. I encourage you to video conference with me over Zoom. I am also happy to schedule an appointment with you.

ABOUT US. We will meet to discuss regression models and their applications from Wednesday, August 24^{th}, 2016 through Thursday, October 10^{th}, 2016. We will meet at the University of San Francisco's downtown campus at 101 Howard Street in room 529. The final exam will be held on Wednesday, October 12^{th}, 2016 from 10.00 to 12.00 in rooms 527 and 529.

ON TAs. There is a TA assigned to assist this class with grading and providing students with guidance. Mrun Bhagwat is a 2015 MSAN alum and. Mrun can be reached via Slack on the main MSAN 601 group channel, as well as in private with direct messages in Slack. Mrun will hold formal office hours every Friday from 10.00 - 12.30 in the 5^{th} floor Agora. Mrun will not respond to email.

ON COMMUNICATION. All formal course material such as the course syllabus, course notes, and data sets will be posted on Canvas. All grades are also posted exclusively on Canvas. All other forms of communication with the instructor will occur through Slack, either in the MSAN 601 group channel, or in private direct messages. You are required to check Slack daily and are responsible for any clarifications, changes and/or updates posted on the MSAN 601 group channel. Emails are discouraged.

ON TEXTBOOKS. There is no formal course textbook required, although acquiring a referential textbook for regression is strongly recommended. Linear regression is oft- and widely-employed across many disciplines, from analytics to medicine to sociology. You will almost certainly run into regression models in your professional careers. This course is custom-designed for the M.Sc. Analytics program at the University of San Francisco, and I have yet to find a singular textbook that treats all of the topics we will discuss in MSAN 601, although there are multiple very good options that will treat upward of 85% of the material. *Applied Linear Regression, 4e* by Sanford Weisberg is an excellent albeit technical treatment of linear regression. The advantage of the aforementioned text is its accompanying text, *An R Companion to Applied Regression, 2e* by Fox & Weisberg, which can be purchased separately. One need not purchase *Applied Linear Regression, 4e* with *An R Companion to Applied Regression, 2e*, as the latter is a standalone text. Note that *An R Companion to Applied Regression, 2e* focuses on the application of regression modeling with R, and lacks a comprehensive mathematical treatment of linear regression. I have found *Applied Linear Regression Models, 4e* by Kutner, Nachtsheim & Neter, or its bigger brother, *Applied Linear Statistical Models, 5e* by Kutner, Nachtsheim, Neter & Li to be both exceptional and very accessible texts. The former (14 chapters) is a less costly, abridged version of the latter (30 chapters). Supplemental chapters in *Applied Linear Statistical Models, 5e* additionally treat the Design and Analysis of Single- and Multi-Factor Studies (11 chapters) as well as Specialized Study Designs (5 chapters), none of which are covered in MSAN 601. Other high-quality texts include *Introductory Econometrics: A Modern Approach, 5e* by Wooldridge, and an *Introductory Econometrics with Applications, 5e* by Ramanathan. Lastly David Dranove, Professor at the Kellogg School of Management at Northwestern University, has a set of ten self-contained lecture notes entitled *Practical Regression.* These lecture notes are accessible, technical and concise, and can be purchased individually (by chapter) for $3.95 from Harvard Business Publishing (`http://hbsp.harvard.edu/`).

ON R. `R` is a powerful open-source programming language and software environment for statistical computing and graphics. The R language is used by many professional statisticians and is making deep inroads in industry as well. R is equipped with a wide variety of statistical and graphical techniques. **The use of R for regression modeling is a course objective, therefore you are not permitted to use any other software for this course**.

ON ATTENDANCE. Formal attendance will not be taken, nor will it be required. You are all graduate students and are expected to mature enough to manage your time intelligently. If you miss lecture(s), you need not explain or excuse yourself formally to me. My objective as course instructor is to ensure that you understand the material to be covered in this course. If you are already familiar with the material or choose to learn it on your own time, that is your prerogative.

ON QUIZZES. Quizzes will be administered every Wednesday from 09.30 - 10.00 for **all** MSAN 601 students.

- Students registered for the MSAN 601 10.00 section will take the quiz at 09.30 in SFH 529
- Students registered for the MSAN 601 13.00 section will take the quiz at 09.30 in SFH 527

ON HOMEWORK. You will be required to complete four homework assignments. They will be posted on the Canvas course webpage. You must work on homework **individually** and **turn in your own individualized write-up**. You may consult with other students in the class regarding homework, but each student should complete all parts of the assignment successfully without assistance. Significant differences between homework scores and test scores may be subject to investigation. Homework is graded as follows:

Description	Grade
All questions attempted and completed with the proper diligence, attention to technical detail, and clarity of presentation.	100%
All questions attempted with minor or serious problems in **one** of the following areas: proper diligence, attention to technical detail, or clarity of presentation.	75%
All questions attempted with minor or serious problems in **more than one** of the following areas: proper diligence, attention to technical detail, or clarity of presentation.	50%
Not all homework questions attempted **OR** not submitted on time.	0%

The above grading scale is designed to encourage students to engage in all homework questions, without fear of reprisal for making mistakes.

ON CASE STUDIES. You will be required to complete two case studies designed to help you synthesize knowledge, suss contextual business issues and reconcile theory with practice. Cases are similar to the types problems you will encounter in your practicums as well as in industry, namely, problems that are ill-defined, incomplete and/or open-ended. I expect case studies to take significantly more time to complete that homework, therefore, you are required to complete case studies in groups of two students. The ability to share and debate ideas with your partner will provide you with a rich environment in which to analyze problems and propose solutions. You may choose your own case study partner or one can be assigned to you by the instructor. You may choose a different case study partner for each case. Each participant in a group is expected to share work equally. You should not allow your name to be placed on a case study deliverable if it does not reflect your own understanding of the material and if you have not made an honest, equitable contribution to the group effort. Students are not permitted to consult with other student groups in the class regarding case studies.

As opposed to homework, case studies are graded on a continuous grading scale. Moreover, they will be competitively graded against other case studies submitted by your peers. This grading scheme encourages students to take deep dives into the data, explore and manipulate the data in creative ways and generate interesting or novel recommendations.

ON DELIVERABLES. All deliverables are subject to the following rules:

1. Submit both an `.R`/`.Rmd` file and a `.pdf` to Canvas for each **homework** submission; no hard-copy is required. Homework may not be emailed to the instructor or to the TA. All homework should be uploaded to Canvas.
2. **Case studies** are to be submitted **both** online to Canvas (only `pdf` file required) as well as in hard copy. Staple all deliverables submitted in hard-copy in the upper left-hand corner and do not include any plastic covers, cover pages, or title pages; simply type your name at the top of each page of your deliverable in a header. Double-sided printing is encouraged.
3. When making reference(s) to *summary* results, include all relevant output in text of the deliverable where it is being discussed, not in an appendix at the back of the deliverable.
4. Do not include a copy of the raw data unless there is a compelling reason.
5. `R` can generate hundreds of graphs and statistical output extremely easily. Only include *relevant* graphs and output in the deliverable. All graphs and statistical output included in the deliverable should be referenced in the text of the deliverable, and should also be indexed, e.g., Figure 1 or Table 2. Everything should be orderly and easy for the grader to read.
6. All deliverables are required to be typed and all graphs and statistical output generated in `R`. Deliverables with *any* handwritten elements will not be accepted and will receive a grade of zero. You are required to submit homework in RMarkdown or LaTeX.
7. All code should be commented in a neat, concise fashion, explaining the objective(s) of individual lines of code.
8. **I will not accept late deliverables under any circumstance.**

ON THE FINAL EXAMINATION. There will be a final written comprehensive examination in this course on Wednesday, October 12^{th}, 2016, from 10.00 - 12.00. The final examination will be a a formal, paper-based conceptual exam. The exam will focus on concepts, i.e., you will not be expected to engage routine calculations, but will rather be expected to know and understand formulae, understand certain relationships and to interpret various regression outputs.

ON GRADING. Part of my job as an instructor is to assign grades fairly and in a manner that reflects the high academic standards at the University of San Francisco and in the MSAN program. Your grade in this course will be computed according to the following weights:

Component	**Weight**
Quizzes	18% ($6 \times 3\%$)
Homework	12% ($4 \times 3\%$)
Case Studies	30% ($2 \times 15\%$)
Final Examination	40%

Final letter grades are assigned on a relative scale. Historically, final grades are approximately distributed as follows:

Letter Grade	A's	B's	C's
Approximate % of Students	20%	65%	15%

This year's grade distribution will necessarily be different. I do not have a predefined letter grade distribution or quota to which I must adhere. The A grade range indicates distinguished performance and competence; the B grade range indicates strong to adequate performance and competence; the C grade range demonstrates weak understanding of the material. A grade of C- is

the minimal passing grade. A grade of F is given for performance that insufficiently demonstrates academic competence.

ON CHEATING. As a Jesuit institution committed to *cura personalis*—the care and education of the whole person—the University of San Francisco has an obligation to embody and foster the values of honesty and integrity. The university upholds standards of honesty and integrity from all members of the academic community, including faculty, students, and staff. All students are expected to know and to adhere to the university's honor code. You can find the full text of the code online at `http://www.usfca.edu/catalog/policies/honor/`. You are also bound by the terms of the MSAN Code of Conduct that you signed prior to matriculating in the analytics program. Refer to ON HOMEWORK and ON CASE STUDIES sections for details regarding student collaboration on each category of deliverable. Plagiarism consists of copying *any* material from *any* source and submitting it as your own original work, regardless of where that material was sourced: the Internet, a book, textbook, or from deliverables perviously submitted by other students. All students involved in any cheating or plagiarized deliverables, i.e., the cheater as well as the person(s) who willfully enabled or facilitated the act of cheating, will be reported to the MSAN Program Director. If you ever have questions about what constitutes plagiarism, cheating, or academic dishonesty in this course, I am happy to discuss with you at your convenience.

ON DISABILITIES. If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF Student Disability Services (SDS) at 415.422.2613 within the first week of class, or immediately upon onset of the disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, please visit `http://www.usfca.edu/sds/` or call 415.422.2613.

ON LAPTOPS. Please bring a laptop to lecture and have `R` installed on it. You will be expected to use `R` in a lecture setting for in-class examples, labs, or exams. I expect you to use your laptops judiciously, refraining from surfing the web or engaging in any other distracting behavior during lecture.