University of South Florida
Muma College of Business
Information Systems and Decision Sciences
ISM 6642 – Statistical Programming for Business Analytics
Spring Semester 2023

Instructor:
Professor Shivendu Shivendu Office: CIS 2079A
Email: shivendu@usf.edu Phone: 813-974-6337 (Not regularly attended)
Office Hours: Tuesday 1:00pm – 2:00pm and by appointment
Web: http://business.usf.edu/faculty/shivendu/

Teaching Associate: TBA

COVID-19 Statement

The health and safety of students, faculty, staff, and visitors on our campuses is our top priority. In response to the current COVID-19 pandemic, the USF community will be working together to support compliance with recommended health and safety standards to optimize the learning experience while minimizing health risks. The Conduct Expectations for all members of the community may be accessed at (Conduct Expected to Support USF Health and Safety Standards).

Additional Covid-19 Procedures

- All students must comply with university policies and posted signs regarding COVID-19 mitigation measures, including wearing face coverings and maintaining social distancing during in-person classes. Failure to do so may result in dismissal from class, referral to the Office of Student Conduct and Ethical Development, and possible removal from campus.

Course Description:

With the proliferation of digital technologies in business, organizations, healthcare systems as well as in social interactions, the ability to understand, analyze, extract insights and interpret micro as well as macro business trends, data-driven insights in various domains including public health, social sciences and healthcare has become increasingly more important today. Widespread adoption of Big Data methods to analyze real time data is making data analytics invaluable for enterprises. This class aims to equip you with high-in-demand in the current job market data analytics as well as empirical analysis skills using SAS.

The course focuses on extracting business insights as well as relevant relationships between variables of interest by leveraging firm's business data, online social media content, as well as data from other

relevant sources for various applications, including (but not limited to) social media analytics, market analysis and demand estimation, customer segmentation, customer relationship management (CRM), and health care management. The class will be hands-on, and the emphasis will be placed on the "know-how" aspect - how to extract and apply data-driven insights to improve business as well as organizational decision making and strategies.

We will analyze real-world data from many enterprises as well as other domains like healthcare and public health using statistical software SAS. This implies that you will learn programming on a widely used statistical analysis software, SAS, along with in-depth applications of advanced statistical methods. The course will also introduce some advanced predictive models with applications in business as well as in healthcare and public health context. Prior programming skill is not required.

Each session of the course will be organized in two modules: Statistical Methods and Tools Module and SAS Programming Module. Some of the course content will be available online on Canvas before the session.

Course Objectives:

- Differentiate, design and assess various data analytics and data mining models
- Identify and translate real-world business problems into data analytics and data mining problems
- Exhibit ability in pre-preparing and visualizing the right data towards these problems
- Develop proficiency in statistical methods
- Implement data analytics models to solve these problems
- Develop proficiency in statistical programming for various business applications
- Enhance knowledge and skills in the current trends in the management and use of BI.

Course Format/Location: In-class Lecture: Tuesday, 8.30 am - 12.15 pm,

Course Materials:

- ***No Textbook for the statistical analysis.***
- ***For SAS Programming, we will use a textbook: A Gentle Introduction to Statistics Using SAS Studio; and a reference book***
- ***A Handbook of Statistical Analyses using SAS by Der and Everitt***

The following books are highly recommended, and the course will cover some of the chapters from some of these books. In addition, relevant articles, reading material will be posted on Canvas.

In this class, we will learn and use SAS Enterprise Guide + and you can access that through your browser from https://usfapps.cloud.com/Citrix/StoreWeb/#/apps/all . Enterprise Guide is a point-and-click, menu- and wizard-driven tool that empowers users to analyze data and publish their results while in background all analysis is done in SAS Base program.

We will borrow many examples, pedagogue, content and material from an excellent book, "A Little SAS Book for Enterprise Guide by Slaughter and Delwiche, 2017 edition". Having access to this book will help

in getting setup and running very quickly. There are good number of free resources available on the Internet to get you started on SAS Enterprise Guide. Another good book is," Data Analysis Using SAS Enterprise Guide by Meyers, Gamst and Guarino. There are many good tutorials available at https://support.sas.com/en/software/enterprise-guide-support.html and other credible sources.

Though the core objective of the class will be using statistical and econometric methods to get meaningful insights from data, we will use SAS as a tool to achieve that objective.

Furthermore, we will also be using material and examples from another excellent source from UCLA: https://stats.idre.ucla.edu/sas/ **.** We will be extensively using material relating to SAS and examples from https://stats.idre.ucla.edu/sas/modules/ as well as from https://stats.idre.ucla.edu/other/mult-pkg/seminars/#SAS . Please take your time to get familiar with the material.

The bulk of SAS documentation is available online, at http://support.sas.com/documentation/onlinedoc/index.html. A catalog of printed documentation available from SAS can be found at http://support.sas.com/publishing/index.html.

1. Introduction to Statistics; http://onlinestatbook.com/Online_Statistics_Education.pdf

2. SAS Programming for R Users;
https://support.sas.com/content/dam/SAS/support/en/books/free-books/sas-programming-for-r-users.pdf

3. Exploring Modern Regression Methods Using SAS;
https://support.sas.com/content/dam/SAS/support/en/books/free-books/exploring-modern-regression-methods-special-collection.pdf

4. An Introduction to Statistical Learning;
http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf

5. Learning SAS by Example; http://onlinestatbook.com/Online_Statistics_Education.pdf

6. An Introduction to the Science of Statistics: From Theory to Implementation;
https://www.math.arizona.edu/~jwatkins/statbook.pdf

7. Advanced Data Analysis from an Elementary Point of View;
https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf

Course Outline: ***Required as well optional readings*** *for each of the classes will be announced and posted.*

Date	Module	Topic
CLASS 1:	**Class 1_Stats_ Module**	**Introduction to Course and overview of statistical concepts** • Probability, statistics, statistical inference, and statistics learning • Common families of distributions • Parametric and non-parametric methods • Data description and simple inference • **Quiz 1: Based on Class 1 Readings**
	Class 1_SAS_ Module	**Introduction to SAS** • basics of using SAS: - SAS environment - program syntax - structure of data, types of data, running program, generating log and output • Chap. 1 of LBS • Assignment 1 posted: Due before class 2
CLASS 2:	**Class 2_Stats_ Module**	**Introduction to Statistical methods** • Random variables and properties of random samples • Sampling, central limit theorem • Statistical testing – Hypothesis test • Interval estimation and interpretation **Quiz 2: Based on Class 2 Readings**
	Class 2_SAS_ Module	**Reading in Data in SAS** • Chap. 2 of LBS • Assignment 2 posted: Due before class 3
CLASS 3:	**Class 3_Stats_ Module**	**Introduction to regression analysis** • Linear Regression Concepts • Assumptions • Coefficient confidence intervals and prediction confidence intervals **Quiz 3: Based on Class 3 Readings**
	Class 3_SAS_ Module	**Working with Data in SAS-I** • Chap 3 and 4 of LBS • Assignment 3 posted: Due before class 4
CLASS 4:	**Class 4_Stats_ Module**	**Regression Analysis-II** • Regression Diagnostics and Advanced Regression Topics • Multicollinearity • Interaction • Model Selection **Quiz 4: Based on Class 4 Readings**

	Class 4_SAS_ Module	**Working with Data in SAS-II** • Chap 5 and 6 of LBS • Assignment 4 posted: Due before class 5
CLASS 5:	**Class 5_Stats_ Module**	**Analysis of Variance** • ANOVA • Between- and Within- Subjects Factors • One factor and multifactor -Between Subjects ANOVA • Within- Subjects ANOVA • Non-Parametric ANOVA **Quiz 5: Based on Class 5 Readings**
	Class 5_SAS_ Module	**Data Visualization in SAS** • Chap 8 of LBS • Assignment 5 posted: Due before class 6
MARCH		**SPRING BREAK**
CLASS 6:	**Class 6_Stats_ Module**	**Categorical Data** • Simple inference for categorical data • Categorical input and categorical output • χ 2 test **Quiz 6: Based on Class 6 Readings**
	Class 6_SAS_ Module	**Statistical Analysis I** • Chap 9 (part) of LBS • Assignment 6 posted: Due before class 7
CLASS 7:	**Class 7_Stats_ Module**	**Non-Parametric Methods** • Overview of non-parametric methods • Kolmogorov-Smirnov test • Wilcoxon's signed-rank test • Mann-Whitney U test • Kruskal-Wallis test **Quiz 7: Based on Class 7 Readings**
	Class 7_SAS_ Module	**Statistical Analysis II** • Chap 9 (part) of LBS • Assignment 7 posted: Due before class 8
CLASS 8:	**Class 8_Stats_ Module**	**Discrete choice models: Binary Choice** • Linear Probability Models • Logit and Probit Models • Hypothesis testing with logit models • Odds ratio • Interpretation of coefficients • Pitfalls and challenges **Quiz 8: Based on Class 8 Readings**

	Class 8_SAS_ Module	**Statistical Analysis III : Logistical Regression** • Chap 18 and 19 of DAU_SAS • Assignment 8 posted: Due before class 9
CLASS 9:	**Class 9_Stats_ Module**	**Discrete choice models: Multiple Choices** • Multinomial Models • Ordinal Logit Models • Censored Regression or Count Data Models: Tobit Models **Quiz 9: Based on Class 9 Readings**
	Class 9_SAS_ Module	**Statistical Analysis IV: Non Parametric Procedures** • Chap 27 and 28 of DAU_SAS • Assignment 9 posted: Due before class 10
CLASS 10:	**Class 10_Stats_ Module**	**Panel Data Analysis** • Fixed effects models • Random effects models **Quiz 10: Based on Class 10 Readings**
	Class 10_SAS_ Module	**Factor Analysis in SAS** • Chap 32 of DAU_SAS • Assignment 10 posted: Due before class 11
CLASS 11:	**Class 11_Stats_ Module**	**Clustering and Classification Methods** • Cluster analysis • Principal Component Analysis • Factor Analysis **Quiz 11: Based on Class 11 Readings**
	Class 11_SAS_ Module	**Exporting Data and Debugging in SAS** • Chap 10 and 11 of LBS • Assignment 11 posted: Due before class 12
CLASS 12:	**Class 12_Stats_ Module**	**Survival analysis and Correspondence analysis** • **Survival function** • **Hazard function** • **Cox's regression** • **Displaying contingency tables and correspondence analysis** **Quiz 12: Based on Class 12 Readings** Assignment 12 posted: Due by EOD next Tuesday
		Project Presentations

A. Project:

This is a group project, and each project will be executed by a two-member team. In special circumstances, a team may have up to three members. In this project, you will use the statistical methods and concepts learned in class to a real-world decision-making situation. You will identify a data-driven decision scenario of business relevance, source appropriate data, clean the data if required, analyze the

data using SAS, and derive actionable insights from your analyses by appropriately interpreting results. To the extent possible you will also do some robustness checks. All data analysis must be executed using SAS.

Your final deliverables include a project report (8-10 pages) and an end-of-semester team presentation. Your project must include the following sections: (1) executive summary; (2) table of contents; (3) problem significance; (4) data source/preparation; (5) hypotheses; (6) descriptive analysis; (7) models including results and interpretation of results; (8) robustness checks; and (9) actionable insights based on the results; (10) limitations of your analysis. Project should not be driven by making predictions that is not the focus of this class.

Feel free to consult with me about your data set or projects. I do give extra-credit to teams who surpass my expectations, experiment with new methods and tools, and help the class learn something new.

A.2 Project Milestones and Deliverables: Important dates related to Project:

Proposal Due on – Start of Class 3
Progress Report Submission/Presentation: Class 9
Project Presentations – Class 12
Final Report Due on – EOD 05/07

A.3 Grading Standards for Group Project: The project will count for 25% of your course grade. For each person the project grade will determined as follows:

Progress Reports and Final Written Report	70%
Final Presentation	20%
Progress Report Presentation	10%

1**Incompletes:** Only in **rare cases**, such as serious illness, will an Incomplete be given. An Incomplete must be requested in writing giving the reason for the request and all appropriate documentation.

Attendance Policy: For a graduate level course, students are expected to take responsibility for full preparation and attendance at every class session. The USF General Attendance policy is at http://www.ugs.usf.edu/policy/GeneralAttendance.pdf

Turnitin Submission Privacy: All student work submitted into Canvas will be subject to plagiarism verification via the *Turnitin* system. In order to comply with privacy laws, students are not required to include personal identifying information, such as your name, in the body of the submitted document. Submitting to the SafeAssign Global Reference Database allows papers from other institutions to be checked against your paper to protect the originality of your work across institutions. Please follow your instructor's instructions carefully regarding what identifying information to include.

Students with Disabilities: Students with disabilities are responsible for registering with Students with Disabilities Services (SDS) in order to receive academic accommodation. SDS encourages students to notify instructors of accommodation needs at least 5 business days prior to needing the accommodation. A letter from SDS must accompany this request. Student responsibilities: http://www.sds.usf.edu

Policy on Religious Holidays: Students who anticipate the necessity of being absent from class due to the observation of a major religious observance must provide notice of the date(s) to the instructor, in writing, by the second class meeting. USF policy on religious days is at http://www.ugs.usf.edu/policy/ReligiousDays.pdf

Academic Integrity: The following USF policies cover student responsibilities for issues of academic integrity:

- *Academic Integrity of Students*: See http://www.ugs.usf.edu/policy/AcademicIntegrityOfStudents.pdf
- *Disruption of the Academic Process:* See http://www.ugs.usf.edu/policy/DisruptionOfAcademicProcess.pdf
- *Student Academic Grievance Procedures:* See http://www.ugs.usf.edu/policy/StudentAcademicGrievanceProcedures.pdf

Student Health and Safety: Educators must report incidents of gender-based crimes including sexual assault, sexual harassment, stalking, dating violence and domestic violence. If a student discloses such in class, in papers, or to an instructor, the instructor is required by law to report the disclosure. The Center for Victim Advocacy and Violence Prevention (813-974-5757) is a confidential resource where you can talk about such situations and receive assistance in confidence. Additional confidential resources on campus are: the Counseling Center (813-974-2831) and Student Health Services (813-974-2331).

Policy on University Closure: In the event of an emergency, it may be necessary for USF to suspend normal operations. During this time, USF may opt to continue delivery of instruction through methods that include but are not limited to: Canvas, Skype, and email messaging and/or an alternate schedule. It's the responsibility of the student to monitor the Canvas site for course specific communication, and the main USF, College, and department websites, emails, and MoBull messages for important general information.

Grading Policy: Following Grading system (weightage) will be used in this course.

In-class Quizzes: 12	18%
Statistical analysis and SAS Programming Weekly Assignments: 12	60%
Group Project and Presentation	22%

Class Readings: All class readings will be posted on Canvas.