MBA 640: Data Analysis for Managers

Moon Oulatta, PhD

Spring, 2024

E-mail: oulat1m@cmich.edu
Office hours: (M/W) 10AM-12:00 PM; (T/Th) 11AM-12PM
Office: Sloan 314
Classroom: Grawn Hall 278
Class time: (T/TH) 16:00-17:55 PM
Phone: 989-774-6460

Course Description

This course exposes MBA students to a rigorous set of empirical methods useful for making data-driven decisions. Here, students will analyze raw data in a rigorous fashion and produce an empirical research idea that can be tested with real data. To achieve the latter, students will mainly focus on **Excel** with optional training available for an open-source statistical software called **R**. Ultimately, the primary objective of this course is to equip MBA students with a robust foundation in statistical theory, empirical modeling, and quantitative skills in order to analyze business data effectively and estimate key empirical models. Consequently, the first part of this course provides a general overview of **Excel**, probability theory, including the basics of estimation theory and statistical inference. In the second part of the course, students learn advanced empirical methods and estimation techniques using **Excel** and **R** (optional).

Prerequisites/Corequisites

Admission to the MBA program.

Required Textbook and Materials

- **Main textbook**: Business Analytics: Data Analysis and Decision Making, 8th Edition
  Author(s): Christian Albright | Wayne Winston
  Textbook ISBN-13:978-0357109953

- **Blackboard**: I will only rely on blackboard to post material pertaining to due dates, exams, grades, problem sets, and study guides.

- **RStudio** (optional): **R** is a programming language and RStudio provides an environment for statistical computing and graphics. First, one needs to install **R** in order to install RStudio. Both are open-source projects (free of charge). RStudio is not required for the course, it is optional, as mentioned in the textbook. However, I will rely on both RStudio and Excel to demonstrate some key statistical concepts used in class. Nonetheless, students are welcomed to use alternative statistical software for their homework (for example, Python, Tableau, Excel, or STATA). The instructions for downloading RStudio are available on Blackboard.

Course Objectives

Upon successful completion of this course, students should be able to:

1. *Understand the fundamentals of probability, estimation, and inference:* we will discuss the properties of random variables and focus on key probability distributions (for example, the normal distribution, t-distributions, chi-square distribution, the f-distribution, and the binomial distribution), sampling theory, statistical inference (hypothesis testing) and confidence intervals.

2. *Manage data effectively:* know how to obtain data from key public and private databases, differentiate between different types of data (for example, nominal versus ordinal) and datasets (for example, cross-sectional data versus panel data), and be capable to work with complex data (for example, import, merge, append, and transform data).

3. *Understand data visualization*: be able to use Excel, Tableau (optional), or **R** (optional) to visualize data by making maps, bar graphs, scatterplots, histograms, and distribution plots.

4. Identity the theoretical underpinning of a *simple linear regression model.*

5. *Estimate a multiple regression model* with both quantitative and qualitative variables.

6. Rely on *advanced statistical methods* (for example, data mining methods, time series models, panel data models, and binary response models) to conduct business research.

7. Develop strong *quantitative skills* using Excel, Tableau (optional), or **R** (optional) to handle large datasets and perform statistical tests.

8. Produce an *empirical research idea* using real data, which can be further developed into a professional working paper.

Course Details

It is important to read the relevant chapters in your book, including the academic articles listed on the syllabus prior to attending class, particularly because the textbook material and the journal articles complement the lectures. Students should be punctual with respect to turning in assignments: I will not adjust grades or provide makeup exams. Exams are assigned on specific dates: I do not allow makeup exams, except for these students whose cases stem from illness

or injury. Given the level, the class content is designed to be rigorous and extremely challenging. Consequently, being on time, attending office hours every week, participating in class, and completing in-class problem sets are all important tasks needed to ensure high performance in this course. This class incorporates an inclusive and equitable environment, which means that students are expected to treat their peers with a high degree of professionalism, kindness, and respect. Additionally, the university's **Title IX** policy encourages access to information and resources to support anyone who has been affected by, knows of, or wants to help prevent an incident or pattern of behavior. Students may contact the Title IX coordinator (**Mary Martinez**) by phone (989-774-3253) to discuss any questions or concerns. Some students may require additional resources to ensure equal access: these students should contact the **accessibility resources center** to discuss ways in which their needs can be met. The accessibility resources center is located in **park library 120**.

Grade Distribution

- **Exams (60%)**: there are two take-home exams (see due dates on class schedule). The content of the exams emanates strictly from lectures and the book.
- **Empirical research idea (20%)**: students should form groups and submit a four-page empirical research idea, testable with real data (see instructions and guidelines on blackboard).
- **Case studies (academic articles review, 20%)**: we will discuss various academic papers to complement the lectures. Students are expected to form groups and select one of the weekly articles listed on blackboard and discuss the latter in class (see case study instructions on blackboard).

Grade	Range (%)
A	94-100
A-	90-93.99
B+	85-89.99
B	80-84.99
B-	77-79.99
C+	73-76.9
C	70-72.99
C-	67-69.99
D+	63-66.99
D	60-62.99
F	0-59.99

Table 1: Grade Distribution (MBA-640)

Schedule and weekly learning goals

The following class schedule is tentative and may change if needed. The weekly topics should mainly serve as a reference for the lectures.

(Week I)

- *Probability review*: in the first week of class, we will briefly review properties of random variables (for example, discrete and continuous random variables), which will then allow us to examine the key probability distributions that we will work with in class (for example, the normal distribution, the standard normal distribution, the chi-square distribution, the f-distribution, the t-distribution, and the binomial distribution).
  - ***Chapter 6: probability and probability distribution***. (required reading)

(Week II)

- *Understanding data*: in the second week of class, we will discuss various types of data (for example, nominal data, ordinal data, etc.), including three different types of datasets (for example, cross-sectional data, panel data, and time series data).
  - *managing data*: I will demonstrate web-scrapping, re-coding, appending, merging, and transforming data using **Excel**.
  - *data visualization*: visualizing data using **excel** and **R** (optional for students).
  - ***Chapter 2: describing the distribution of a variable***. (required reading)
  - ***Chapter 3: finding relationships among variables***. (required reading)

(Week III)

- *Sampling and estimation theory*: we will discuss the characteristics of an **estimator**, the **central limit theorem**, the **sampling distribution** of means and proportions, including **confidence intervals**.
  - *In-class example*: here, I will provide a practical example that illustrates the sampling distribution of means using **R** (optional for students).
  - *Introduction to **sampling design***: purpose of sampling, sampling errors, bias, and three different sampling methods (for example, **random sampling**, **stratified sampling**, and **cluster sampling**).
  - *Hypothesis testing*: understanding the concept of hypothesis testing, performing **one-sided** and **two-sided** tests. We will go over the **two-sample t-test**, the **f-test** for equality of variances, and the analysis of variance **(ANOVA) test**.
  - ***Guess speaker (`Tanima Ahmed`, Economist, Gender Group, World Bank): Understanding Survey Design***
  - ***Chapter 8: statistical inference***. (required reading)
  - *Take-home exam # 1 due on (02-04-24)*

(Week IV)

- *The simple linear regression model*: we will derive the **ordinary least squares (OLS)** estimator for a simple linear regression model and discuss the underlying classical assumptions of the OLS estimator.

- *simple linear regression model (selected issues)*: we will discuss the limitations of a simple linear regression model and how to address those limitations.
- *data workshop*: we will estimate a linear regression model in **Excel** with real data. We will examine the relationship between **nominal variables** by using the **chi-square test of independence**, including all appropriate graphs. We will also examine the relationship between **continuous variables** by relying on **covariance analysis** and the appropriate **bivariate graphs** (for example, scatter plots). Lastly, we will rely on the **gamma test of association** to examine the relationship between **ordinal variables**.
- ***Chapter 9-4: simple linear regression***. (required reading)

(Week V)

- *The multiple regression model*: we will discuss the underlying classical assumptions of a multiple linear regression model.
  - *least absolute deviations (LAD) estimator*: we will discuss an alternative estimator to the OLS estimator (for example, we will discuss the LAD estimator).
  - *multiple regression with qualitative variables*: we will discuss the main differences between the **linear probability model** (LPM), the **logistic** regression, and the **probit** model.
  - *data workshop*: we will estimate various multiple regression models using real data and perform diagnostic tests in Excel.
  - ***Munnell, Alicia, Tootell, Geoffrey, Browne, Lynn E. and McEneaney, James, (1996), Mortgage Lending in Boston: Interpreting HMDA Data, American Economic Review, 86, issue 1, p. 25-53.*** (article review)
  - ***Chapter 9-5: multiple regression***. (required reading)
  - ***Chapter 9-6: modeling possibilities***. (required reading)
  - ***Chapter 9-7: validation of the fit***. (required reading)
  - ***Chapter 10: regression analysis, statistical inference***. (required reading)
  - ***Chapter 16-2b: logistic regression***. (required reading)
  - *Exam # 2 take-home due (02-18-24)*

(Week VI)

- *Data mining (basic intro)*: businesses rely on data mining to understand consumer preferences and make data-driven decisions. Here, we will apply data mining methods (for example, **clustering** and **text mining**) to examine **structured** and **unstructured data**.
  - ***Sunil Erevelles, Nobuyuki Fukawa, Linda Swayne, Big Data consumer analytics and the transformation of marketing, Journal of Business Research, Volume 69, Issue 2, 2016, Pages 897-904***. (article review)
  - ***Chapter 17: data mining, clustering, and market based analysis***. (required reading)
- *Introduction to time series analysis*: we will review the properties of time series data, perform seasonal adjustment of data, and go over different unit-root tests.

- *forecasting with ARIMA models*: we will rely on autoregressive processes to perform out-of-sample forecast of time series data.
- *forecasting using a vector autoregressive model (VAR)*: I will use a simple bivariate VAR model to analyze and forecast two endogenous variables.
- ***Fattah J, Ezzine L, Aman Z, El Moussami H, Lachhab A. Forecasting of demand using ARIMA model. International Journal of Engineering Business Management. 2018;10. doi:10.1177/1847979018808673*** (article review)
- ***Chapter 11: time series analysis and forecasting***. (required reading)

(Week VII-IIX)

- *Panel data regressions*: we will discuss the underlying assumptions of a panel data regression model. We will examine the main differences between the fixed-effects estimator, the random-effects estimator, and the first-difference estimator.
  - *research day (1)*: we will use the remainder of class time to work on the research papers.
  - *research day (2)*: we will use the remainder of class time to work on research papers.
  - *research idea presentations*: students will present the main findings of their four-page research idea in class on February 29th.[1]
  - ***Chen, Yinghong and Hammes, Klaus, Capital Structure Theories and Empirical Results- a Panel Data Analysis.*** (article review)

Academic Integrity

Refer to the University's **academic policies and guidelines** for information pertaining to academic integrity and honesty policy. Students are responsible to adhere to the honor code in all academic endeavors.

[1]The final draft for the group research paper is due on the last day of class, which is on March 9th.