454 A – Statistical Experiments and Causal Inference

Please see RAPS for classroom and meeting schedule.

Course Description

At a high level, this course will teach to you ask yourself the following question:

"What's the counterfactual?"

When attempting to measure the impact of any "treatment" X (e.g., a marketing action like an ad campaign) on an outcome Y (e.g., sales), the temptation is often to compare today to yesterday. We ask "Were our sales with the new ad campaign higher than our sales used to be with the old ad campaign?" But this is never the right question to ask, as many things change over time. For example, perhaps demand for our product is seasonal, or perhaps a competitor has entered the market since we began the new campaign. The right question to ask is "Were our sales with the new ad campaign higher than they would have been if we continued the old ad campaign, all else equal?"

If we want to understand the *causal* effect of X on Y (e.g., sales), we can—sometimes—run an experiment. In such settings, we can construct "treatment" and "control" groups that are identical save for the treatment we apply. Three sessions in this course give you experience with designing and measuring the results of experiments.

When running an experiment isn't possible or cost effective, we can attempt to estimate the impact of X on Y using statistical techniques on historical data. The difference between the scenario we observe (today's sales with the new campaign) and the one we do not (today's sales with the old campaign—an unobservable counterfactual scenario) is the true impact of the new campaign on our sales. Two sessions in this course train you to think in terms of counterfactuals, and to run analyses in ways that better estimate the *causal* impact of X on Y when working with historical data.

Requirements

Prerequisites:	Stats Core (203A) or equivalent
Software:	Microsoft Excel (required)
Required Text:	Mastering 'Metrics by Angrist & Pischke
Optional Texts:	Statistics for Management and Economics by Gerald & Keller (203A text)

Difficulty Level

This is not an introductory course – it is an **intermediate-to-advanced** course that presumes solid familiarity with statistics and some experience working with data. Having a strong understanding of linear regression by the end of session two is necessary. This is easier if you have already taken 203B, but not strictly required. If you received a B- or lower in the stats core, I would **strongly recommend** you take 203B before taking this course.

Course and Assignment Schedule

This course has little in the way of reading. In lieu of that, it is heavier on post-class homework. However, each assignment is typically started in the final hour of class as a group. One assignment is an individual assignment to ensure each student develops the basic statistical skills for the course. Three others are group assignments in which students will work together to apply those skills to a real-world case.

The first (individual) assignment will follow the more traditional format of statistics course homework, favoring repetition and commitment of technique to memory. Group cases, by contrast, are not like those in traditional stats courses. Because the primary objective of this course is to improve your thought process for tackling analytical questions, little emphasis is placed on memorization of substantive learnings; rather, emphasis is placed on developing your ability to think about data. This means that I don't expect you to be able to answer all case assignment questions perfectly on a first pass, because the value of the assignment is not retrieving a fact from memory, but the experience of trying to think through the problem.

Assignments are always due at 8 am the day of class. See the above schedule for specific due dates for each assignment. Assignments should be submitted on Canvas in Word or PDF format, with all statistical output, tables, and figures included either in the appropriate part of your write-up, or in an appendix and referenced in your write-up.

Weekly Schedule: Topic, Readings, and Assignments

Week	Topic	Reading Due	Assignment Due
1	A/B Testing, Part 1	MM Intro, Ch. 1	n/a
2	A/B Testing, Part 2	n/a	HW 1: A/B Testing
3	Difference-in-Differences	MM Ch. 2 & 5	Team Case 1: A/B Testing
4	Fixed Effect Models	n/a	Team Case 2: Dif-in-Dif
5	Field Experiment Case	Case Prompt	Team Case 3: Fixed Effect models

Grading

Each assignment is worth 25% of your grade. Attendance and participation are a part of your grade, but their effect is non-additive: Missing one class will cap your grade at 90% (if you score a 93%, your grade will be reduced to a 90%). Missing a second will cap your grade at 70% (likely in the range of C- to C+, depending on the curve). You cannot pass the course if you miss three classes, as you will have missed more than half the material we cover. Absences can be excused under certain circumstances (e.g., serious illness or the death of a family member), in which case you will not lose points.

Note that GSM events such as Big Bang are announced sufficiently far in advance that you will know whether or not this course conflicts with events of interest to you. Absences for GSM events will therefore not be excused.

Team Formation

You will have the same team for each case. Students will self-select who they wish to work with; no one will be forced to work with anyone else. Teams should consist of no more than three students. You may work alone if you wish or if you cannot find a team. As a rule, we do not assign teams in this course, as the risk for free-riding is high in challenging courses. No one will ever be required to work with anyone else against their will.

Course Policies

Anti-free-riding policy

The intention of this policy is to prevent free-riders from taking advantage of their team members. I want to **protect you** from others who might take a spot on your team and never show up, hoping you'll take the path of least resistance and do the work without them. Such instances are rare, but sadly do happen from time to time. The policy is as follows:

1) Teams may fire team members if they are not pulling their weight. Fired team members must work alone on all assignments.

2) Teams will be required to grade each member of the team at the end of the year.

3) If I determine a student was free-riding on an assignment after they were reported for doing so, they will receive a zero on that assignment.

These policies are not aimed at students who may struggle with the material. **If you're still a little uncomfortable with statistics, do not let this discourage you**. I am always happy to help you with the assignments, and your classmates are generally happy to help those who make a genuine effort.

Laptop policy

Laptops will be needed for some in-class exercises. However, when not doing those exercises, laptops must be closed. In the past I have had students request to have them open to take notes, but observational data tells me that laptops are more often a distraction than a note-taking device.

UC Davis and GSM Policies

Notice of the Code of Academic Conduct

Students are expected to conform with the code of academic conduct, which can be found here:

http://sja.ucdavis.edu/files/cac.pdf

Academic Affairs, as well as the Academic Director of the MBA Program, will be notified of any violations, and will take appropriate action.

Cross-Attendance Policy

Intention to take a make-up course in a location other than the one in which you are enrolled now follows a specific procedure. A request to attend must be made to your professor, with a justification provided.

If this request is made at least one week in advance of the section you wish to attend, and you will be attending in either San Ramon or Sacramento, you should also CC the Instructional Operations contact for the location and get both their and the professor's permission. No Instructional Operations contact need be contacted for make-up courses in Davis, only your professor.

If the request is made less than one week in advance of the section you wish to attend, you must also CC the Senior Assistant Dean of Student Affairs, who must give first approval before the professor and Instructional Operations contact can approve.

Relationship to other courses

There are several courses pertaining to marketing analytics (and business analytics more generally) at the GSM. The faculty make an effort to ensure each is its own unique offering, without much redundancy. Below I briefly identify some (but not all) of the courses someone interested in the 254 sequence might take, how those courses relate to the 254 courses (and each other), and how they differ.

0. Primary Statistics Courses

203A – Analysis for Managers ("Stats 1")

203B – Forecasting and Managerial Research Methods ("Stats 2")

285 – Times Series Analysis and Forecasting ("Stats 3")

1. Methodologically-Focused Courses

249 – Marketing Research

269 – Business Intelligence Technologies and Data Mining

254A – Causality and Statistical Experiments

254B – Marketing Analytics

2. Objective-Focused Courses

234 – Pricing

239 – Digital Marketing

243 – Customer Relationship Management

248 – Marketing Strategies

282 – Supply Chain Management

I separate courses other than the primary statistics courses (203A, 203B, 285) into two groups: (1) courses that have a broader approach pertaining to statistical methodology, and (2) courses that teach analytics pertaining to a specific objective (or set of related objectives) in depth. The first set of courses are more general and teach you how to think like a data scientist, while the second set are more likely to be directly related to specific job functions or a specific class of decisions that marketing managers need to make. For example, Pricing is included in the second bucket, as it tackles pricing-related analytics in depth.

The Marketing Analytics (254) sequence is in bucket (1) and is most closely related to Marketing Research and Business Intelligence Technologies and Data Mining. These courses cover both statistical methodology and data, with Marketing Research focusing more on the collection and analysis of primary data and BIT&DM focusing more on analysis of secondary data (especially "big data").

254A is a blend of data collection (via experiments) and statistical methodology under the methodological umbrella of causal inference. 254B is focuses primarily on statistical methodology for secondary data, like BIT&DM, but not in the realm of big data. 254B focuses on (a) how to recognize what questions a dataset allows you to answer and what it does not, and (b) how to answer a question of interest with that dataset.