Statistics By Default **2026**

Week 0: Syllabus

Author: Andrew Lizarraga

Welcome

- The aim of this (unofficial) course is to expose UCLA students to real-world statistics.
- We cover topics in marketing, finance, health, academia, ect.
- Importantly, we will use the theory you learned from class in practical and compelling ways.
- Warning: I'm assuming you have some basic knowledge of Python.

To preface this: I've been growing increasingly concerned about the UCLA undergraduate population in terms of their general knowledge (I don't mean GPA), their willingness to help and cooperate with each other, and their overall motivation.

As some of you already know, the job market has become increasingly difficult due to economic circumstances beyond your control. During these times, I've found that students will often settle for what they can get (which is fine, by the way: take a job now to support yourself in the heat of the moment.) But I want to actively combat this tendency.

The tools and skills that students learn from the stats department are actually exceptionally valuable for job prospects. However, these tools are often not well motivated, or any motivation comes from toy examples instead of real-world ones. Moreover, the lack of natural cooperation prevents the student population from lifting each other up. I find it especially sad that I don't constantly hear stories of UCLA alumni reaching back out to the undergraduate population to help them career-wise.

The only way I can imagine fixing this is through a culture change. So I'm proposing this course as a low-stakes opportunity to expose you to topics you've likely never seen. The topics are challenging, which is also why they represent employable skills. I'm also hoping this encourages natural collaboration as you tackle them together.

List Of Topics

1. A/B Tests: How businesses use them incorrectly and lose money without realizing.
2. Mean Absolute Deviation (MAD) porfolio optimization; viewing optimizaion as risk reduction.
3. Statistical arbitrage of cryptocurrency markets and why it's so difficult.
4. Python Package Management
5. Python Coding Algorithms (Recursion)
6. SQL Coding Heuristics (Lookups, Joins, ect.)

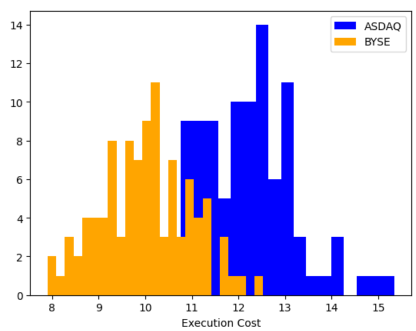
The plan is that I'll host these sessions on zoom, every **Friday: 11:00am (Beginning Jan 9th, 2026)**:

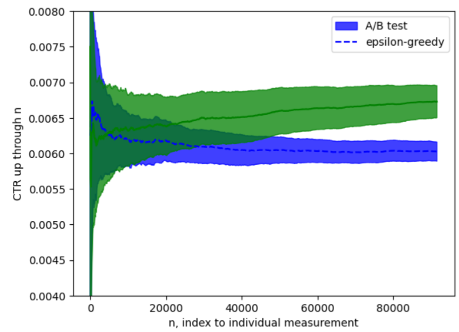
`https://ucla.zoom.us/j/*******777?pwd=YmR6VDFQaEk4Myt4ZUF0WStpVTJaZz09`

To join you need to email me (with your ucla email) at: **andrewlizarraga at ucla dot edu**. Please let me know what you want out of the course. Below is a taste of the first 3 topics. Depending on pacing and feedback from the students, we'll consider moving into topics 4 through 7 or amend some of them from the syllabus.

0.1 AB Tests

We are asked to compare two stock exchanges: ASDAQ and BYSE. We hypothesize that BYSE has a cheaper exchange rate on average. We took 100 samples from each exchange and compre their exchange costs. What can go wrong in concluding that BYSE is cheaper on average? Let's say there is no issue, and that infact BYSE is better. This means that since we needed to purchase stocks on ASDAQ, we had an opportunity loss for not putting that money into BYSE. Is there a way to perform this A/B test make the correct decision that BYSE is better and minimize the opportunity cost (the regret) of spending on ASDAQ (which is required to run the test).





0.2 Mean Absolute Deviation Portfolios

Naievely, some may think that when assembling a portfolio (a stock portfolio), that you want to maximize your returns. A slightly less naieve approach is to include some sense of risk (risky things tend to be more profitable, but they also tend to be catastrophic). However the dual to this is to make a portfoio that minimizes risk subject to some desirable lower-bound on returns (r).

But what if we don't have too much liquid cash to purchase stocks (our weights w). Can we also enforce our model to give the consumer a reasonable finite selection of stocks (w)? One approach is to take the above formulation and use a mean absolute deviation (MAD) loss function:

$$MAD(w) = \frac{1}{T}\sum_{t=1}^{T}|\sum_{j=1}^{J} w_j(r_{t,j} - \bar{r}_j)|$$

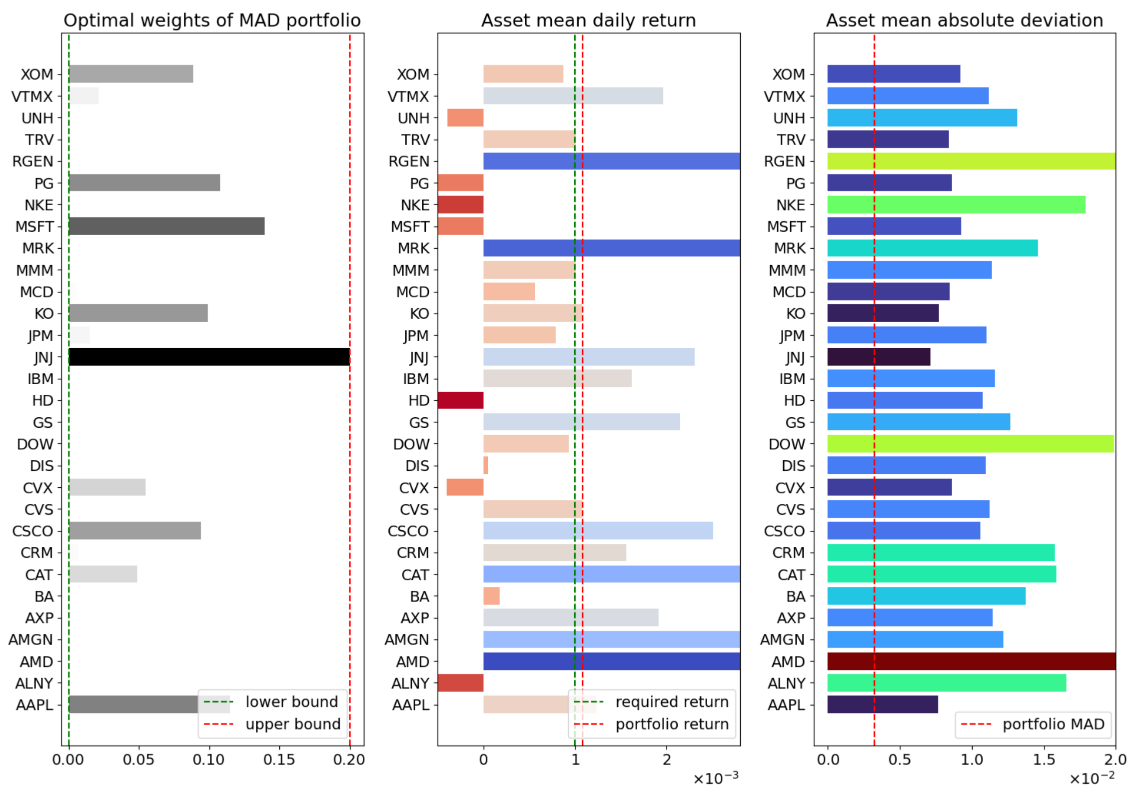


Figure 0.1: (Left): The proportion of each stock you should purchase. (Middle) the loss or gain of each asset over a period of interest; the green line is the desired return of the portfolio; the red line is the actual return of the portfolio. (Right) The MAD (i.e. the risk) of each asset. Cold blues mean low risk (also their lengh suggest this). greens transitioning to dark reds are higher risk.

0.3 Statistical Arbitrage

Suppose we want to find price differences across exchanges. The graph of all possible ways you can trade an exchange money for a small subset of cryptocurrencys can be seen in the figure below. Image we consider a few assets of interest and the trades to other assets we can make.

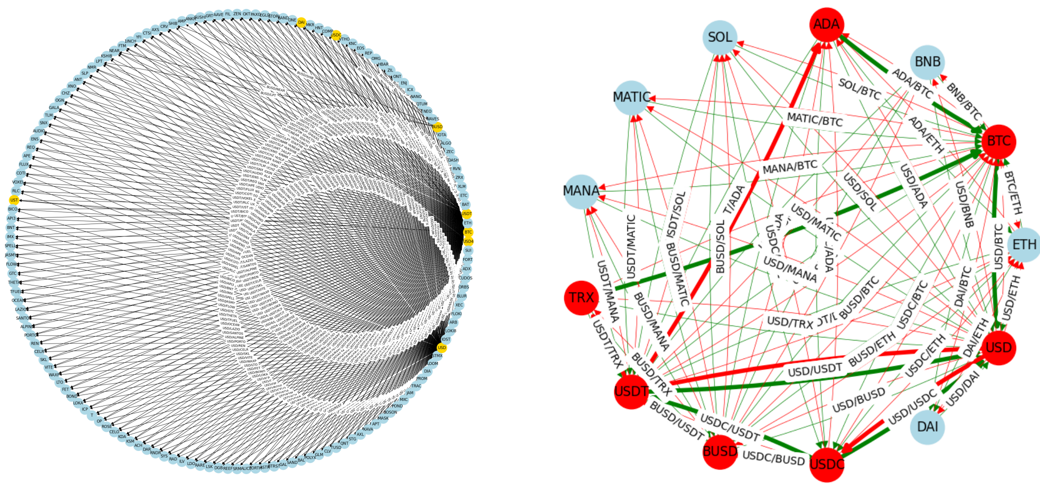


Figure 0.2: Even if we perform a single transaction, there are $n-1$ other possible tranactions we can make. As you can imagine the number of possibilities explodes if we consider taking more actions. How can we handle this? The sub-graph on the right is an optimal solution of the exchanges you should trade on. Green means a gain, red means a loss, and the thickness of the arrow is proportional to the amount of how much volume should be traded on that exchange.