

# The Road To Machine Learning and AI

David Rios

Brownstone Investment Group

# Topics we will over

- Summarizing Data
- From Data to Populations
  - Why Probability ( Cardano)
  - Two great ideas in statistics : confidence interval / hypothesis test
- Does X explain Y : Classical Statistics
  - ANOVA / Linear Regression
- How Machine Learnings helps
- How AI Helps

# Summarizing Data

- With numbers
  - Mean / Median / Mode -- Outliers vs non-outliers.
  - Standard deviations / IQR
  - Correlation (to measure interactions)
    - Standard deviations / IQR
- With Charts
  - Histograms
  - Box Plots (to measure interactions)
  - Scatter plots (to measure interactions)

# From Data to Populations







Our goal is to extend our past observations to future predictions.

Idea 1 : Confidence Interval – Report a range, not a point, make a market on the range. “At what price”

Idea 2 : Hypothesis test – Is the data consistent with previous beliefs. “What are my stop outs”

We need to math of probability to accomplish this. Remember the job of math is to say, if you believe A then B is either logically consistent or inconsistent with that belief.

Table 1: PMF for both a fair and a loaded die.

$\Omega$						
Fair	1/6	1/6	1/6	1/6	1/6	1/6
Loaded	1/2	1/10	1/10	1/10	1/10	1/10

## Basic Probability

- Starting from the above assumptions how do we find the probability of winning a round of craps.

# The Goal of Statistics : Does X Explain Y?

We care about statistics because there is some random number, the price of gold, that we want to be able to predict. The difference between statistics and physics is physics wants to predict it with 100% accuracy and we are ok with less than 100%.

# Linear Regression Example

Mean of CMT 10yr : 4.1

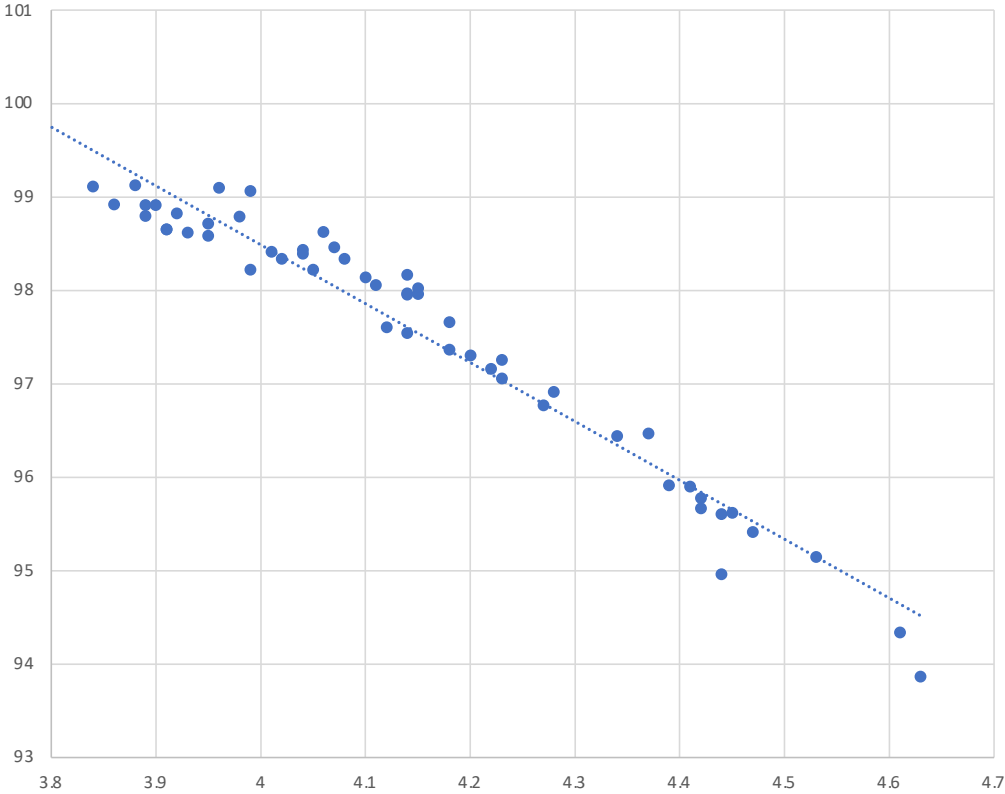
Mean of FNCL 5.0 : 97.6

Std Dev of CMT : 20bp

Std Dev of FNCL 5.0 1.4

Correlation = - 98%

Notice : We can quantify the effect of X on Y, this is our hedge ratio



# Machine Learning

- Classical Statistics has a hard time with non-linear data and with weird interactions between terms
  - Non Linear Data
  - Weird interactions are interactions without the traditional form of  $X * Y$ , i.e. sometimes X and Y interact, sometimes they don't. Does X explain Y : Classical Statistics
- How Machine Learning works. With the power of computers we can divide our data set into many sub clusters and perform the analysis on each cluster. The key is the computer decides how to sub cluster the data. Because there are any possibilities high computing power is needed.
- The downside of machine learning is the models lack 'explainability' its very hard to say if we change the first input by 10 how much will Y change unless we rerun the model.



# Artificial Intelligence

With equations and algorithms its hard to draw the line between AI and ML. In popular terms its easy

Machine Learning : Numbers go in, Numbers come out

Artificial Intelligence : Words go in, words come out

Before AI computers have to be given very specific instructions in their own programming language. The goal of AI is to interpret English and convert it to those specific commands.