Date.....

Lesson 10: Sampling Distributions & the Central Limit Theorem:

1. Probability to Statistics

→ This begin a set of lessons that will be more data oriented in the way that you are applying ideas, & less probability oriented.

→ Often we use statistics to verify conclusions of probability through simulation.

2. Sampling Distribution

→ The distribution of a statistic.

3. Notation for parameters vs statistics

	Parameters	Statistic
Mean	μ	$\bar{x}$, $\hat{\mu}$
Standard deviation	σ	s, $\hat{\sigma}$
variance	σ^2	s^2, σ^2
Proportion	π	p, $\hat{\pi}$
Regression coefficient	β	b, $\hat{\beta}$

Date.....

* Two important mathematical theorems for working with sample distributions include:
  1. Law of large numbers
  2. Central limit theorem.

1. Law of large numbers
   → It says that as our sample size increases, the sample mean gets closer to the population mean.

   → But how did we determine that the sample mean would estimate a population mean in the first place? How would we identify another relationship b/w parameter & statistic like this in future?

   Three of the most common ways are with the following estimation techniques:
   1. Maximum likelihood estimation.
   2. Method of moments estimation.
   3. Bayesian Estimation.

2. Central limit Theorem
   → The central limit theorem states that with a large ~~sam~~ enough sample size the sampling distribution of the mean will be normally distributed.
   → The central limit theorem actually applies for these well known statistics:
   1. Sample means ($\bar{x}$)
   2. Sample proportions (p)
   3. Difference in sample means ($\hat{x}_1 - \hat{x}_2$)
   4. Difference in sample proportions ($p_1 - p_2$)

16/05/2020 Date.....

Bootstrapping

→ Bootstrapping is sampling with replacement.
→ Using random.choice in python actually samples in this way, where the probability of any number in our set stays the same regardless of how many times it has chosen.
→ Flipping a coin & rolling a die are kind of like bootstrap sampling. as well, as rolling at 6 in one scenario doesn't mean that 6 is less likely later.
→ Bootstrapping is used in leading machine learning algorithms:
  - Random forest
  - Stochastic gradient boosting

How to take boot sample?

```
bootsamp = coffee_red.sample(200, replace=True)
```