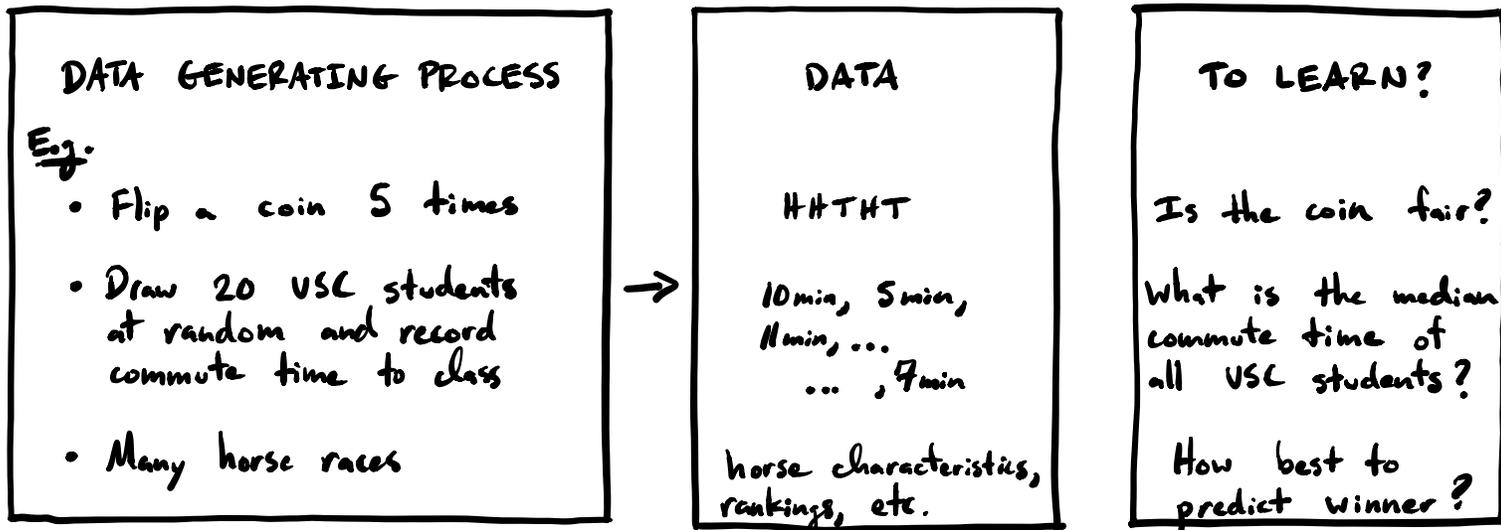
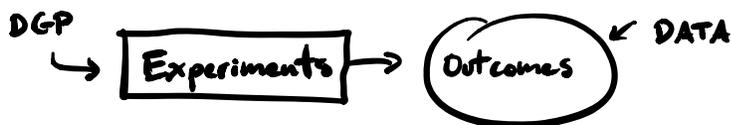


# OVERVIEW OF STAT 512

Goal of Statistics: In statistics we study what can be learned from random outcomes (data) about the process which generates them.



We may often regard the DATA GENERATING PROCESS (DGP) as a sequence of experiments, and the DATA as the sequence of outcomes.



STAT/MATH 511 began with the definition of an experiment:

Defn: An experiment is a process which generates an outcome such that

- (i) more than one outcome is possible.
- (ii) the possible outcomes are known.
- (iii) it is not known in advance which outcome will occur.

Then the following were discussed:

- Set theory - to work with collections of possible outcomes of an experiment
- Probability theory - to be able to assign probabilities to sets of outcomes of an experiment

- Random variables - to work with numerical encodings of experimental outcomes.

- \* Probability distributions
- \* Cumulative distribution function (cdf)
- \* Probability density/mass function (pdf/pmf)
- \* Expected value and variance
- \* Many commonly encountered pdfs/pmts
- \* Quantiles
- \* Moments and moment generating functions (mgfs)
- \* Joint and marginal distributions when two or more random variables are considered
- \* Conditional distributions and conditional expectation
- \* Independence of random variables
- \* Covariance and correlation
- \* Hierarchical models

... So we have not yet learned how to learn from the DATA about the DATA GENERATING PROCESS.

In STAT 512, we begin to learn how to learn from DATA.

... In these steps:

- I. We need some more tools for working with random variables, namely transformations of random variables. This will feel like STAT/MATH 511 continued.
- II. Then we focus on a DATA GENERATING PROCESS called a random sample. A random sample is the collection of outcomes from  $n \geq 1$  identical experiments conducted independently, encoded in random variables,  $X_1, \dots, X_n$ , say, such that  $X_1, \dots, X_n$  are independent and all have the same distribution.

III. We will spend much time studying sampling distributions—the distribution of some function  $T = T(X_1, \dots, X_n)$  of the values in a random sample is called the sampling distribution of  $T$ . We will, for one, give ample consideration to the function

$$T(X_1, \dots, X_n) = n^{-1} \sum_{i=1}^n X_i,$$

so that  $T$  is the sample mean, and we will study its distribution under different settings.

IV. We then turn to learning about the distribution from which the values  $X_1, \dots, X_n$  were generated. Our "learning" will happen through

- (a) Estimation of quantities related to the unknown distribution
- (b) Inference about these quantities via
  - Confidence Intervals
  - Tests of hypotheses (this rather in STAT 513)

What we will have learned about sampling distributions will play a key role in (a) and (b).