Statistical Estimation & Inference
The Signal vs. the Noise

As Nate Silver* will tell you, the difficulty in statistical inference is separating the signal from the noise.

- If we flip a fair coin, it is possible that the outcome will be a sequence of all heads, just due to random chance.
- If this is the only sample that we see, then how can we separate the signal (that the coin is fair, say) from the noise (the sequence of all heads that we observe).
Paul the Octopus

- A German animal oracle who successfully predicted the outcome of all 7 of Germany’s matches in the 2010 World Cup
- He also predicted Spain to win the Cup final
- How likely is his success rate?

> dbinom(7, 7, 1/2) # (½) ** 7

[1] 0.0078125

- Other factors: He chose Germany 11/14 times, Spain (twice), and Serbia once. Similar flags. Likely color blind, but perhaps has preference for horizontal shapes.
A statistical model is a set of assumptions about how sample data are generated, characterized by parameters (e.g., $\mu$, $\beta_0$, and $\beta_1$).

Examples:
- Here’s a simple one: $Y = \mu + \epsilon$, where $\mu$ is a model parameter representing the mean of a population, and $\epsilon$ is a random error term (a.k.a. noise).
- Here’s another one: $Y = \beta_0 + \beta_1 X + \epsilon$, where $\epsilon$ represents noise/error.
Statistical Estimation & Inference

- Design and build a parametric model intended to explain data
  - about a population, or a sample of a population
- Estimate its parameters
- Use the model to draw inferences
  - about a population
- Quantify the uncertainty in those inferences

“All models are wrong, but some are useful.” -- George Box
Population

- A population is a set of individuals under study, chosen by the statistician, or nowadays, the data scientist.
- When the entirety of the population is observable, we can compute descriptive statistics like population mean, population variance, histograms, etc.
- But typically, the entirety of the population is unobservable.
Sample

- A sample is an observed subset of the population
- A statistician’s goal is to study the sample, and from it, draw inferences about the entirety of the population

Polling is a typical application of statistical inference:
- For example, ask a few college students their favorite ice cream flavor; then infer the favorite flavor of all students.
Statistical Machine Learning
Supervised Learning (a.k.a. Function Approximation)

- Given a set of labelled training examples, \( \{(x_i, y_i) \mid i \in 1, \ldots, n\} \), learn a function \( f: X \rightarrow Y \)

- Typical goal is to learn a function \( f \) from features \( X \) to labels \( Y \) that
  - Either predicts well on test data, or
  - Explains the dependence of \( Y \) on \( X \)
Statistical (Supervised) Machine Learning

- Given a set of labelled training examples, \( \{(x_i, y_i) \mid i \in 1, \ldots, n\} \), learn a conditional probability distribution \( P(Y \mid X) \)
  - E.g., the probability of playing tennis, given that it is hot, sunny, not too humid, and not too windy outside

- Two key approaches
  - Discriminative models: Learn \( P(Y \mid X) \) directly
    - E.g., Linear and logistic regression
  - Generative models: Learn \( P(X \mid Y) \), and apply Bayes’ rule to infer \( P(Y \mid X) \)
    - E.g., Naive Bayes