STAT 135 Lecture 1

Henry Liev

24 June 2025

Definition 0.1
What is **statistics**?
Statistical Inference: Study of answering questions and making decisions with data. Uses a lot of probability.
Probability vs. Statistics:
Probability: You start with a model that you know. Have a biased coin that lands heads with probability $\frac{3}{4}$. We toss it 10 times, what is $P(H = 6)$?
Statistics: You start with data. Flipped it 10 times, got 6 heads.

Example 0.2
What is a good guess for the probability of heads? (Estimation)
Is the coin fair? (Hypothesis testing)
What is a reasonable range of values for p? Good real questions but fuzzy. Statistics is not just math. How do we frame the problem clearly so you can answer mathematically?

Example 0.3
Real applications:
Is this vaccine effective? Safe? (Clinical trials)
Who will win the election? (Polling)
Should OpenAI launch new agent feature? (A/B testing)

Definition 0.4
Population:
True state of the world. Probability distribution.
Probability distribution:

- $x_1, x_2, \ldots, x_n$ heights of everyone in the US $\frac{1}{N}$ on each person
- Coin: Heads (p), Tails (1-p)

Definition 0.5
Parameters:
Property of the population that you are interested in.

- Mean
- Spread
  - Variance
  - Standard deviation
- 75th percentile

Definition 0.6
Sample aka data:

- Draws from populations.
  $X_1, X_2, \ldots, X_n \leftarrow$ often independent
- Assume draws are with replacement
- Samples are random variables

Example 0.7
Flip a possibly biased coin 5 times, get 3 heads. What is a 'good guess' for p? $\frac{3}{5}$? $\frac{2}{3}$? What if we had 0 heads?

Definition 0.8
Estimation:

1. Construction: Come up with a way to estimate a parameter
2. Evaluation: How good is our estimator? Bias? Concentrated around the true value of the parameter θ? Variance? Consistency?
3. Optimality Theory: Can we come up with a best estimator?

Example 0.9
Example of Estimators:

- Heights in the US: $x_1, x_2, \ldots, x_n \overset{\text{iid}}{\sim}$ population of heights.
  - $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

  Sample $X_1, X_2, \ldots, X_n$.
  Sample Mean $= \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Possibly biased coin n flips, X heads.
  $\hat{p} = \frac{X}{n}$ (Sample Fraction)

Definition 0.10

Estimators are random variables. Concentrated around a truth θ, tighter distribution = better estimator. We can study the properties of estimators as random variables.

Example 0.11

Expectation of $\bar{X}$
Population with mean μ
$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$
$E(\bar{x}) \stackrel{?}{=} \mu \rightarrow E[\frac{1}{n}\sum_{i=1}^{n} x_i] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$
$x_1, x_2, \cdots x_n$ independent

Example 0.12

Variance of $\bar{X}$
$X_1, X_2, \ldots, X_n$ independent, mean μ, variance σ^2
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$

Takeaways:

1. Statistics = working backwards from data to make model. (Inference)
2. Parameter = Feature of population
3. Estimator = Function of the data which si a guess at parameter
4. Estimators are random variablesCan study distributions of estimators to understand them