Statistical Inference for Estimation

UC Boulder Foundation for Data Science Specialization

October 27, 2024

Contents

Bilal Naseem

1 Introduction to Inference

A *Poll* is when a person goes out and randomly select people from a large population to form a sample, and study that sample to figure out what the larger population wants.

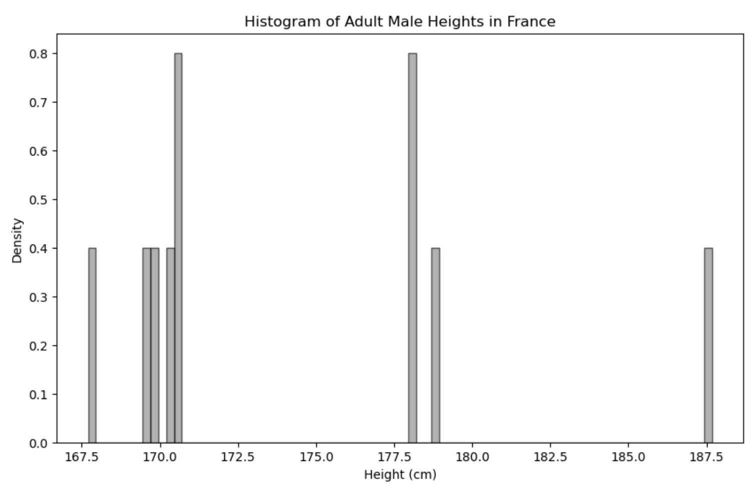
Statistical Inference

Drawing conclusions about a larger population based on a relatively small sample.

Point Estimation: Estimating a quantity of interest in the larger population.

What is the average height of all adult males in France?

In order to estimate the average height of all adult males in France, we first took a sample of 10 adult males and measured their heights. Below is a histogram of the heights of 10 adult males in France. The x axis is the height in cm in bins and the y axis is the density. Density is a number such that the total area of each bar is going to represent the proportion of the people in our sample that we saw in the corresponding bin.

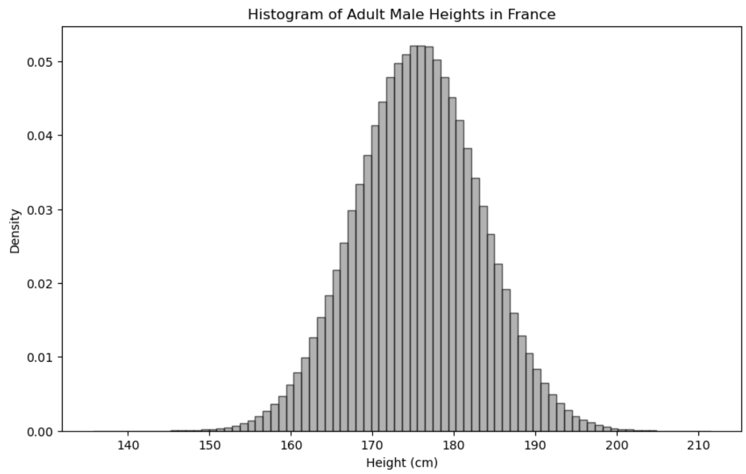


Density: It is a number such that the total area of each bar is going to represent the proportion of the people in our sample that we saw in the corresponding bin.

For 1 million samples, the Histogram is going to take the shape of a bell curve. **This represents perfect information** i.e. the sample is representative of the population or a model for the idealized population.

1.1 Population Mean μ and Sample Mean $\bar{X}$

μ is the average in an ideal population. It is no the sample average.



Because of the symmetry of the Bell curve, the true average height μ is going to be the peak of the histogram.

If i take many samples of size 100 and compute the sample mean $\bar{X}$, they are going to pile up and make a **sampling distribution**. We use the sampling distribution to estimate the true mean μ.

2 Review of Random Variables

2.1 Discrete Random Variables

Let X be a random variable. Before a coin is flipped, the value of X is unknown. It's waiting on the results of the experiment involving randomness.

Random Variables

A random variable is a mapping from the set of outcomes of an experiment involving probability to the set of real numbers.

It is convention to use capital letters when talking about random variables.

$$X = \begin{cases} 1, & \text{if Heads} \\ 0, & \text{if Tails} \end{cases} \tag{1}$$

Since random variables talk about probability, we can say that:

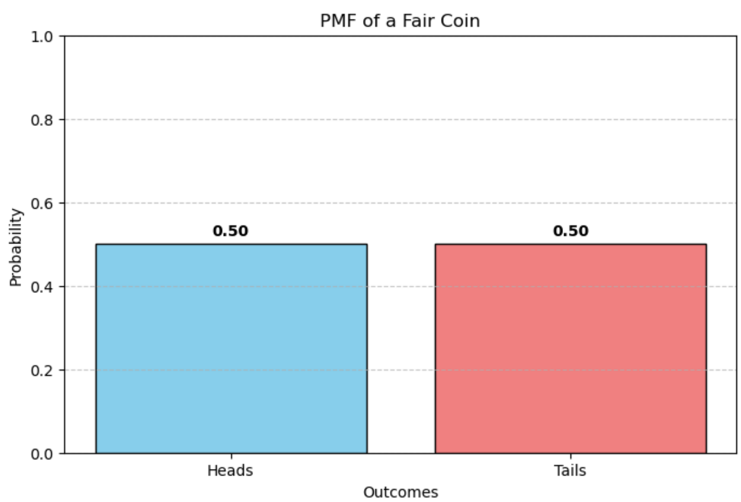
$$X = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases} \tag{2}$$

And a better way to write this would be as a function:

$$f(x) = P(X = x) = \begin{cases} \frac{1}{2}, & \text{if } x = 0 \\ \frac{1}{2}, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

This is known as a probability mass function (pmf) and is denoted by $f(x)$. When we have multiple random variables, we use subscripts to distinguish them for e.g. $f_X(x)$, $f_Y(y)$.

The pmf tells us how the probabilities of different outcomes of a random variable are distributed. The below image shows the pmf of a fair coin. Since we are plotting density, the total area of each bar is the probability of that outcome.



2.1.1 Bernoulli Distribution

We can generalize that if the probability of getting heads is p and the probability of getting tails is $q = 1 - p$, the pmf becomes:

$$f(x) = P(X = x) = \begin{cases} 1 - p, & \text{if } x = 0 \\ p, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

This model depends on the parameter p. This random variable is so common, it gets a name i.e. **Bernoulli Random Variable or Distribution with parameter** p.

We can write any distribution in short hand as follows:

$$X \sim \text{Bernoulli}(p) \tag{5}$$

2.1.2 Indicator function

Let A be a set of real numbers. The Indicator function indicates if a number is in the set or not.

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

with this Indicator function, we can say that X is a random variable that takes on only two values, 0 and 1. With this Indicator function we can write the Bernoulli probability mass function (pmf) as:

$$P(X = x) = p^x (1 - p)^{1-x} \cdot I_{\{0,1\}}(x) \tag{7}$$

We re-wrote the pmf as equation 7 because it is more concise as compared to equation 4. If x is not in set $\{0, 1\}$, the pmf is 0. Indicator function are also helpful in proving proofs.

2.1.3 Geometric Distribution

The geometric distribution models the number of trials needed to get the first success in a sequence of independent Bernoulli trials with a success probability p. In other words, if X represents the number of trials until the first success, then X follows a geometric distribution with parameter p.

$$X \sim \text{Geometric}(p) \tag{8}$$

Conditions:

- Repeated independent Bernoulli trials

- Only 2 outcomes: 0 and 1

Derivation of the Probability Mass Function (PMF)

Considering $P(X = 4)$. For this, we need our first success to be on the 4th trial. We need **F**, **F**, **F**, and then **S**.

$$P(X = 4) = (1 - p) \cdot (1 - p) \cdot (1 - p) \cdot p = (1 - p)^3 p$$

General Formula of the Geometric Distribution

In general, we can say that the first success is on the X^{th} trial, meaning we will have $X - 1$ failures. So we can write the general formula of the geometric distribution as:

$$\begin{aligned} P(X = x) &= (1 - p)^{x-1} p && \text{for } x = 1, 2, 3, \ldots \\ P(X = x) &= 0 && \text{for other values of } x \end{aligned}$$

We can use the Indicator function to write as a single general formula"

$$P(X = x) = (1 - p)^{x-1} p \cdot I_{\{1,2,3,\ldots\}}(x) \tag{9}$$

Plot of the Geometric Distribution

The minimum value of X is 1 and there is no maximum value. The mode is always 1. As we move towards the right, the probabilities decrease.

Important: incomplete: exponentiol, binomial etc. graphs

2.2 Continuous Random Variables

sd