

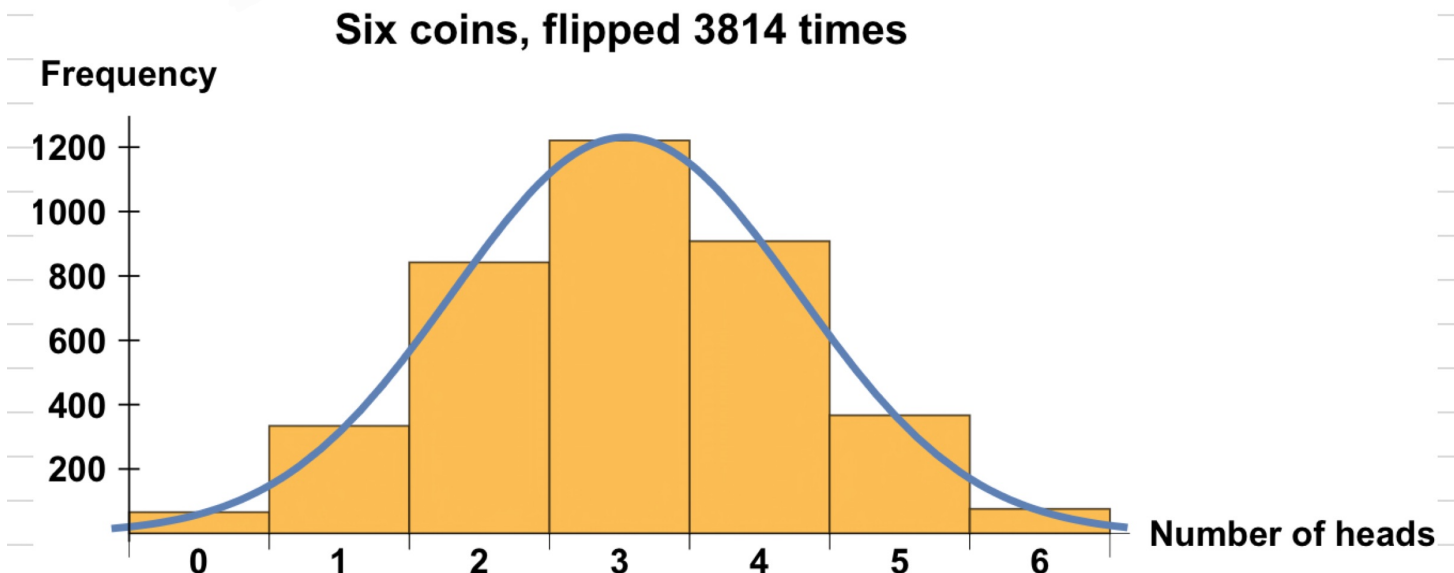
Statistics.

Consider the experiment of flipping six coins and recording the number of heads.

A 2018 TV Calculus class did this 3814 times. Results:

# of heads	Frequency
0	66
1	334
2	842
3	1221
4	908
5	367
6	76

Here's a histogram:

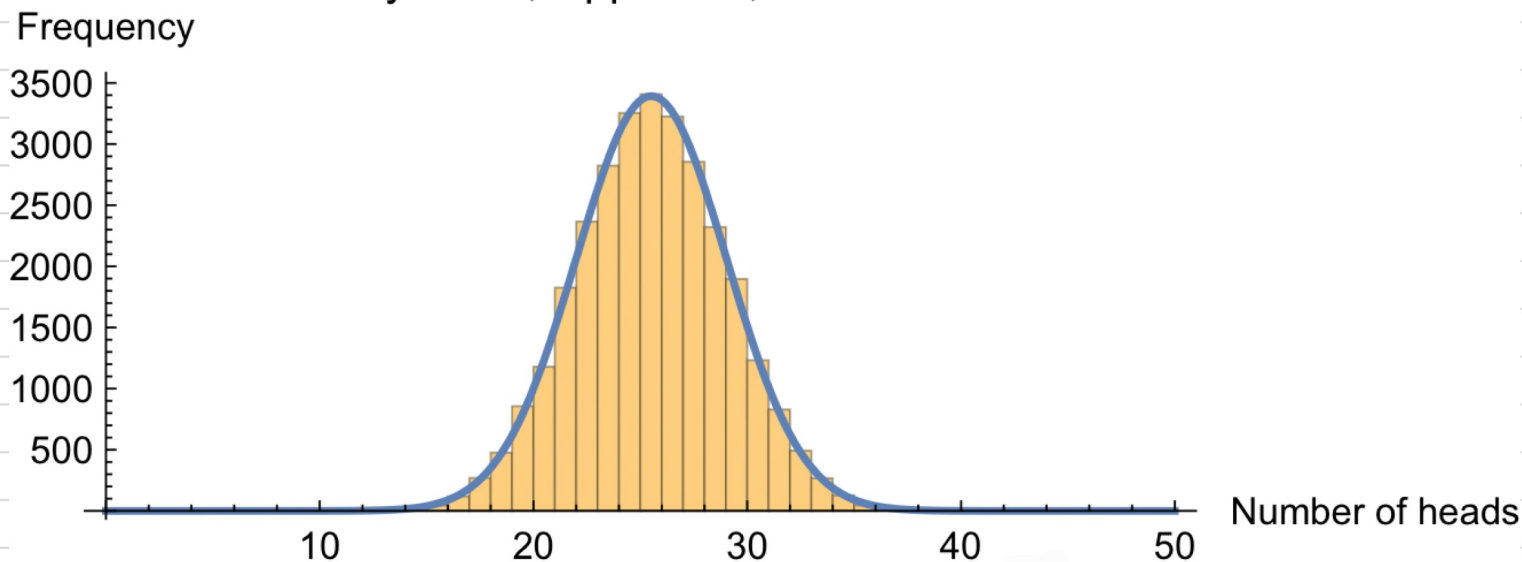


Note the (rough) bell shape: a certain "normal curve" (sketched on the histogram) fits the data fairly well. But which normal curve (and what's a normal curve)? Answers soon.

2

Next, we simulated flipping 50 coins, recording the # of heads, and repeating 30,000 times. Results:

Fifty coins, flipped 30,000 times



A normal curve fits the data quite closely.

This illustrates a central result in probability:

The Central Limit Theorem (CLT).

If each trial of an experiment comprises many small, independent factors, all of which behave similarly, and many trials are performed, then the outcomes of the experiment will follow a roughly normal distribution.

[Proof omitted.]

Interlude: some formulas.

Consider a data set

$$X = \{x_1, x_2, \dots, x_n\} \text{ of real numbers.}$$

We define the mean \bar{x} and standard deviation s (std dev) of the data by:

(A) General formulas.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \left\{ \begin{array}{l} \text{measures "central tendency"} \\ \text{of the data} \end{array} \right.$$

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}} \quad \left\{ \begin{array}{l} \text{measures} \\ \text{"spread" of} \\ \text{the data.} \end{array} \right.$$

(B) Formulas for grouped data.

If the data takes only the distinct values y_1, y_2, \dots, y_k , and the value y_i happens f_i times for each i , then

$$\bar{x} = \frac{f_1 y_1 + f_2 y_2 + \dots + f_k y_k}{n},$$

$$s = \sqrt{\frac{f_1 (y_1 - \bar{x})^2 + f_2 (y_2 - \bar{x})^2 + \dots + f_k (y_k - \bar{x})^2}{n-1}}$$

(These are different formulas for the same quantities as in part (A).)

Example:

for our "six coins, flipped 3814 times" data, we have

$$\bar{x} = \frac{66 \cdot 0 + 334 \cdot 1 + \dots + 76 \cdot 6}{3814} = 3.042;$$

$$s = \sqrt{\frac{66(0 - \bar{x})^2 + \dots + 76(6 - \bar{x})^2}{3813}} = 1.236.$$

Note: we compute that
 $\bar{x} - 3s = -0.666$
 $\bar{x} + 3s = 6.75.$

So: all possible data values $0, 1, 2, \dots, 6$ lie in the interval $[\bar{x} - 3s, \bar{x} + 3s]$. That is, all data is "within three standard deviations of the mean." This exemplifies the "empirical rule." More on this later.