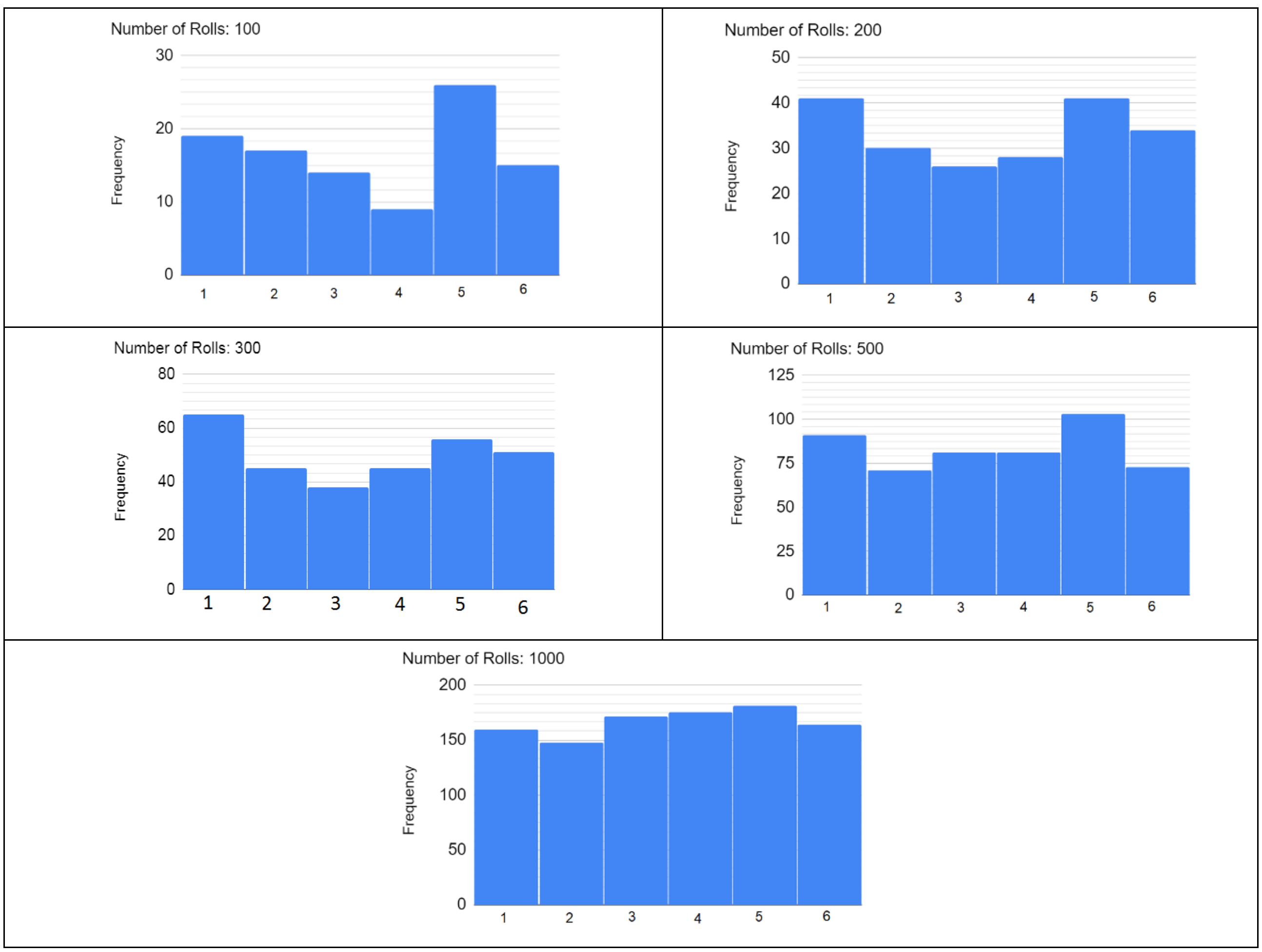
HISTOGRAMS: UNCERTAINTY
STUDENT PAGES

DESCRIPTION
Scientists use histograms to analyze data. In particular, particle physicists rely on histograms when looking for rare events hidden inside “big data.” Sometimes the probability of a particular interaction occurring is small. In such cases, particle physicists collect huge amounts of data in the hope of finding this interaction as a small bump in the histogram. In this activity, students will learn how particle physicists determine the uncertainty in their results.

<u>Part 1</u>
What do we know?
- Histograms are graphs that show how often data values fall within a particular bin.
- The vertical axis of a histogram is always frequency.
- The intervals on the horizontal axis are called bins.

What do we need?
- Sample histograms for rolling dice
  (Students rolled six-sided dice and recorded the number for each roll. The following histograms show the results after 100 rolls, 200 rolls, 300 rolls, 500 rolls and 1,000 rolls.)



Histograms generated using https://academo.org/demos/dice-roll-statistics/

What do we do?

- **Answer** the following questions about the histograms:
  1. What number occurs most often after 100 rolls?
  2. What number occurs least often after 100 rolls?
  3. After 200 rolls, do the same numbers occur most often and least often?
- **Describe** the likelihood of rolling a particular number after 1,000 rolls.
- **Determine** if the resulting histogram is identical for every 100 rolls.
- **Make a claim** about the likelihood of rolling a particular number supported by evidence and reasoning.

Part 2

What do we know?

- Histograms are graphs that show how often data values fall within a particular bin.
- The vertical axis of a histogram is always frequency.
- The intervals on the horizontal axis are called bins.
- When data is collected by counting, the underlying statistical distribution is a Poisson distribution because counting can never be negative. Since particle physicists perform counting experiments, the uncertainty in a count is determined using the standard deviation for the Poisson distribution

$$\pm\sqrt{N}.$$

  Since our histogram represents counting data and can never have a negative bin height, each bin has an uncertainty of $\pm\sqrt{N}$.

What do we need?

- **Scenario:**
  Muons emerge at a particle angle phi (φ) after collisions inside the ATLAS detector. Is there a preferred angle for muons emerging from the collision? The angle phi (φ) is measured around the LHC beam pipe which is the direction of motion for the incoming particles. Figure 1 shows how to measure the angles viewing down the beam line.

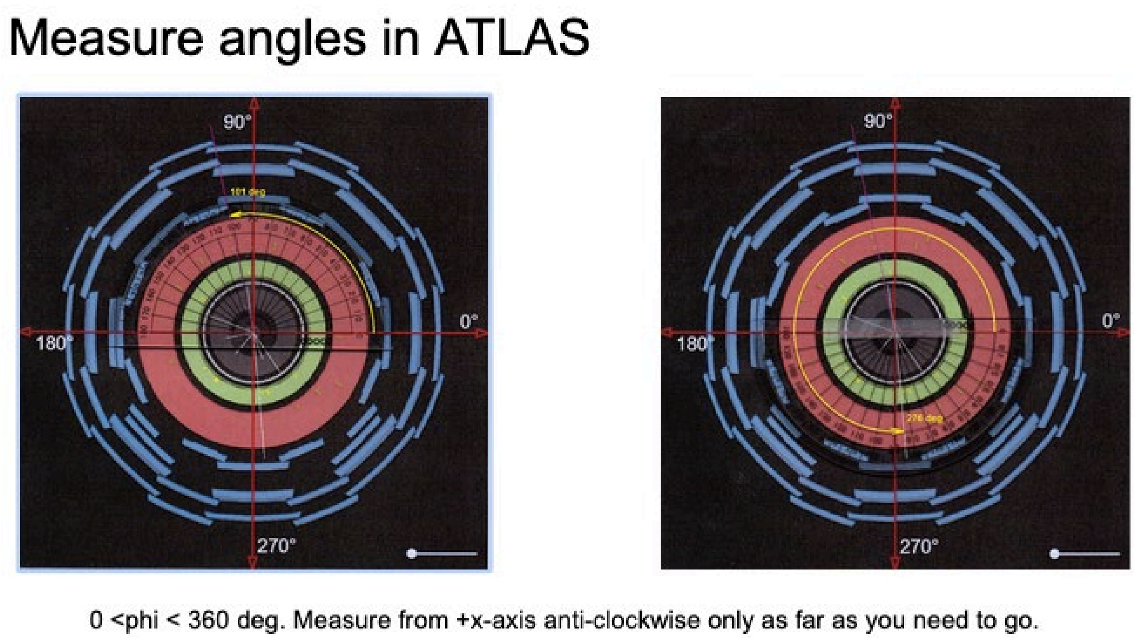


Figure 1: Method for measuring angle φ in the ATLAS detector viewing down the beam line.

Figure 2 below shows the results of measuring thousands of events.

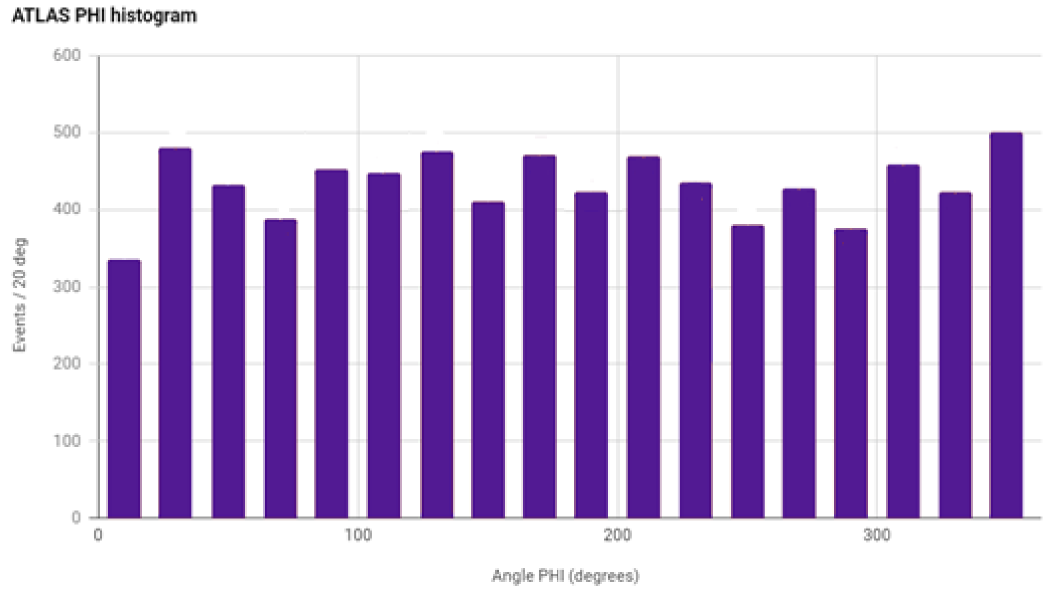


Figure 2: Histogram of the angle muons exit the collision.

What do we do?

- **Calculate** the uncertainty for each bin.
- **Construct** error bars of the correct size for each bin.
- **Make a claim** about whether there is a preferred angle for muons emerging from the collisions. Support your claim with evidence and reasoning.

<u>Part 3</u>

What do we know?

- Histograms are graphs that show how often data values fall within a particular bin.
- The vertical axis of a histogram is always frequency.
- The intervals on the horizontal axis are called bins.
- In a Gaussian (or normal) distribution, the data values can be less than or greater than the mean. Figure 3 below shows the shape of a typical Gaussian distribution. The standard deviation in the Gaussian distribution is approximated by finding width of the Gaussian when the frequency is half of the maximum value and dividing by 2 (Full-Width Half-Maximum over 2, or FWHM/2).

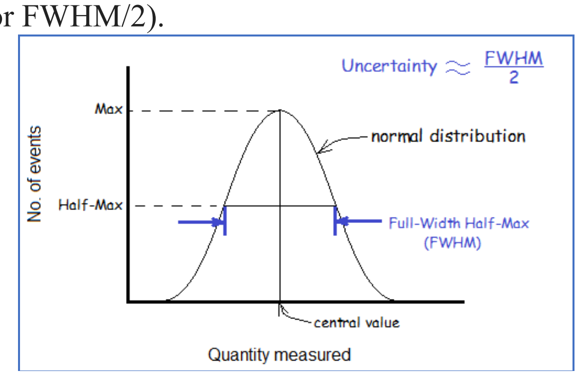


Figure 3: Defining uncertainty for a Gaussian distribution.

What do we need?

- Histogram of data collected by rolling marbles to try to knock hidden marbles out of the way. Each student made 10 tries and recorded the number of hits.

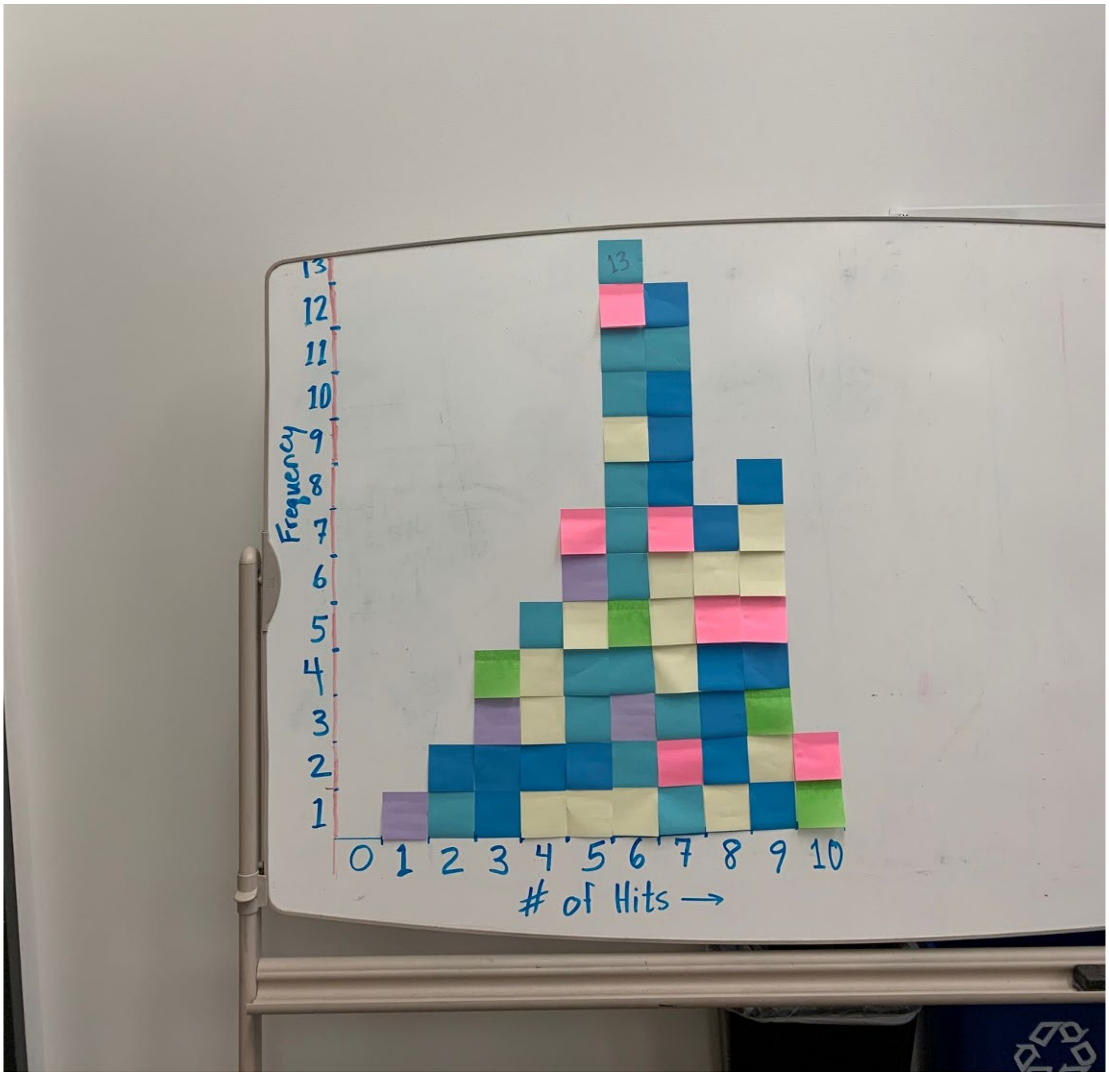


Figure 4: Histogram of *Rolling with Rutherford* data.

What do we do?

- **Calculate** the uncertainty of each bin.
- **Draw** the error bars on each bin.
- **Determine** the uncertainty of the peak using FWHM.
- **Make a claim** about the most likely number of hits including uncertainty. Support your claim with evidence and reasoning.