

HISTOGRAMS: UNCERTAINTY

TEACHER NOTES

DESCRIPTION

Scientists use histograms to analyze data. In particular, particle physicists rely on histograms to find new particles and to measure the characteristics of particles. 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. This activity builds histogram skills required in many of the other activities in the Data Activities Portfolio. In this activity, students will construct histograms, identify the best value to represent the data, and report the uncertainty in their answers.

STANDARDS ADDRESSED

Next Generation Science Standards

Science and Engineering Practices

- 4. Analyzing and interpreting data
- 5. Using mathematics and computational thinking

Crosscutting Concepts

- 3. Scale, proportion, and quantity

Common Core Literacy Standards

Reading

- 9-12.4 Determine the meaning of symbols, key terms . . .
- 9-12.7 Translate quantitative or technical information . . .

Common Core Mathematics Standards

MP2. Reason abstractly and quantitatively.

AP Physics 1: Algebra-Based and AP Physics 2: Algebra-Based Science Practices

Science Practice 4

The student can plan and implement data collection strategies in relation to a particular scientific question.

Science Practice 5

The student can perform data analysis and evaluation of evidence.

IB Physics

Topic 1: Measurement and Uncertainties

- 1.2.6 Describe and give examples of random and systematic errors.
- 1.2.7 Distinguish between precision and accuracy.
- 1.2.8 Explain how the effects of random errors may be reduced.
- 1.2.11 Determine the uncertainties in results.

ENDURING UNDERSTANDINGS

Scientists must account for uncertainty in measurements when reporting results.

LEARNING OBJECTIVES

Students will know and be able to:

- Determine the uncertainty for a particular bin.
- Determine the best value to represent the data.
- Determine the uncertainty in the best value to represent the data.

PRIOR KNOWLEDGE

Students should be able to:

- Collect and organize data.
- Explain the difference between representations of central value: mean, median and mode.
- Construct a histogram given a set of data. This information is covered in the portfolio activity *Histograms: The Basics*.

RESOURCES/MATERIALS

- Ruler
- Calculator
- The importance of histograms for CERN data analysis:
<https://home.cern/news/news/computing/big-data-takes-root>
- Drawing a histogram by hand:
<https://www.youtube.com/watch?v=EqI HVMTaPiA>
Note that the presenter uses the term *classes* while particle physicists use the term **bins**.
- The wiki page for histograms:
<https://en.m.wikipedia.org/wiki/Histogram>

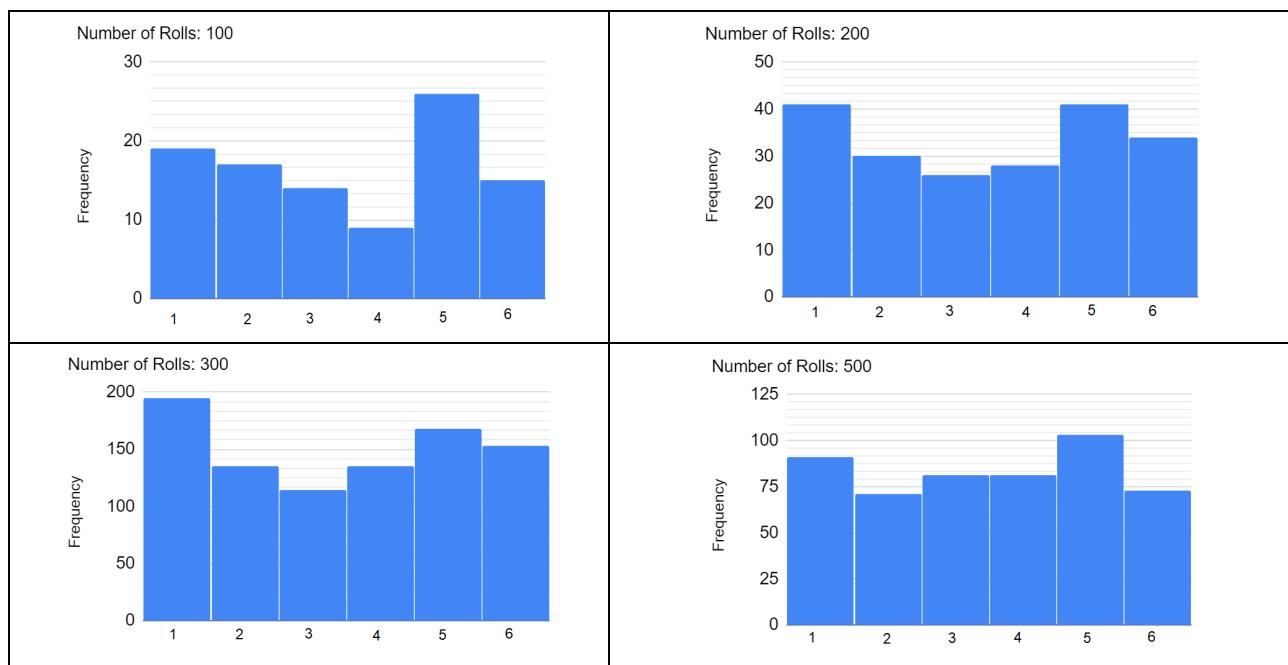
BACKGROUND MATERIAL

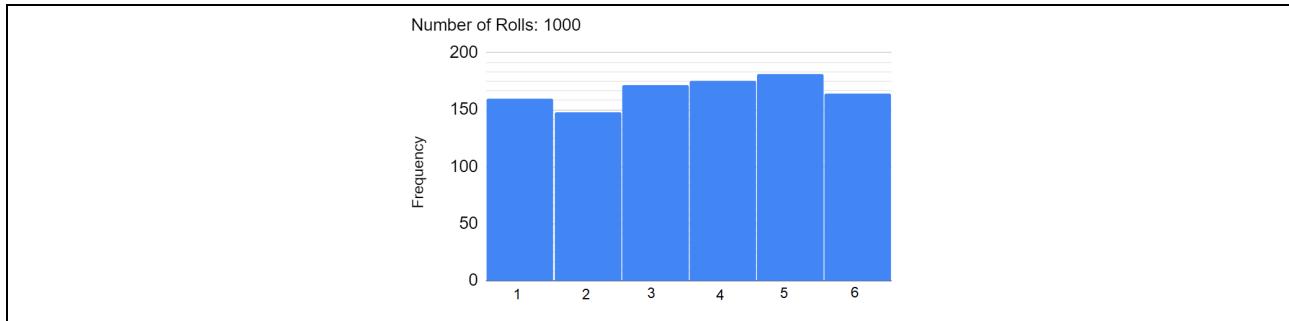
For a review of the methods for constructing histograms, refer to the portfolio activity *Histograms: The Basics*.

IMPLEMENTATION

Part 1:

Imagine that you rolled dice over and over again. You would not expect to get the same number with each roll, but with a large number of rolls, you would expect to have rolled the same number of ones, the same number of twos, etc. The student pages present the following histograms and ask questions to lead them to see that there is no preferred number when rolling a six-sided die. Also, they should notice that each set of 100 rolls yields a different result, but if the number of rolls is large, then there seems to be no preferred number. When all of the sides have equal probability of being rolled, then the die is said to be a fair die.





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

Part 2:

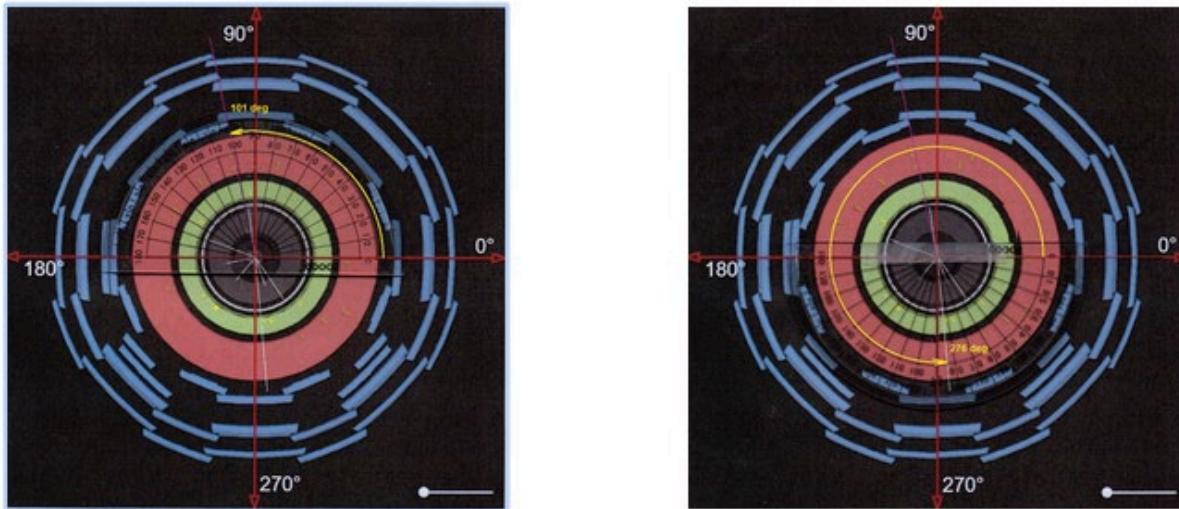
When collecting data by counting, the underlying statistical distribution is a Poisson distribution because counting can never be negative. Since particle physicists perform counting experiments, the uncertainty in a count is determined using the standard deviation for the Poisson distribution

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

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

Muons emerge at a particle angle phi (ϕ) after collision events inside the ATLAS detector. The physicists wanted to see if there was a preferred angle for the muons to emerge from a 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 angle.

Measure angles in ATLAS



$0 < \phi < 360$ deg. Measure from +x-axis anti-clockwise only as far as you need to go.

Figure 1: Method for measuring angle ϕ in the ATLAS detector.

The yellow angle line in the left image shows an angle $\phi = 101^\circ$. The yellow angle line in the right image shows an angle $\phi = 276^\circ$. These data are sorted in 20-degree bins based on the angles around the LHC beam pipe. Figure 2 shows the results of measuring thousands of events. When

each bin is given its error bar, the bins are all basically the same height indicating that there is no preferred angle φ for the muons to emerge from collisions.

The student pages have this histogram without error bars. Their task is to determine the size of the error bar for each bin and correctly draw the error bar in each bin.

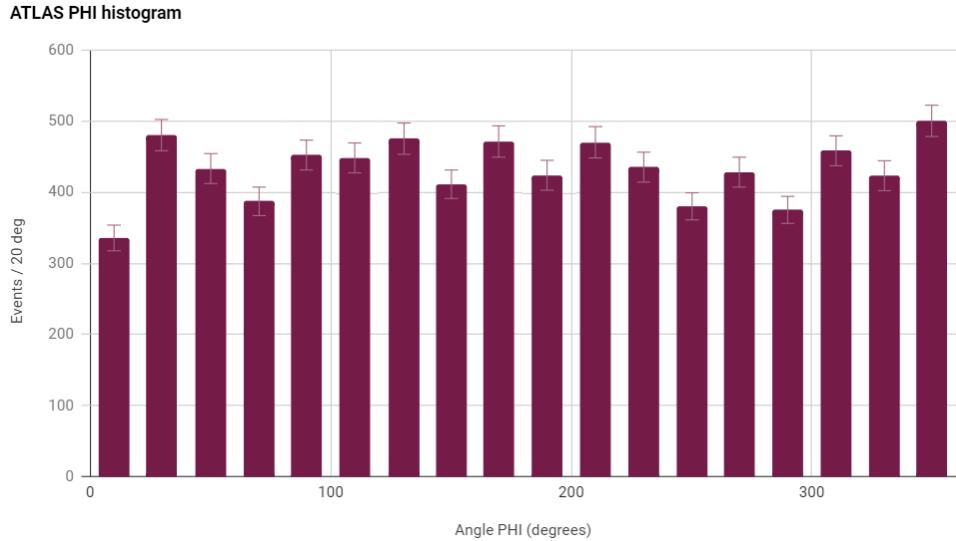


Figure 2: Histogram of the ATLAS results for the distribution of angle φ .

We get pretty much the same result in each bin at around 450 counts but not always the same result, even though the number should not depend at all on the value of phi (φ). This is due to natural statistical fluctuation.

Part 3:

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 the 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).

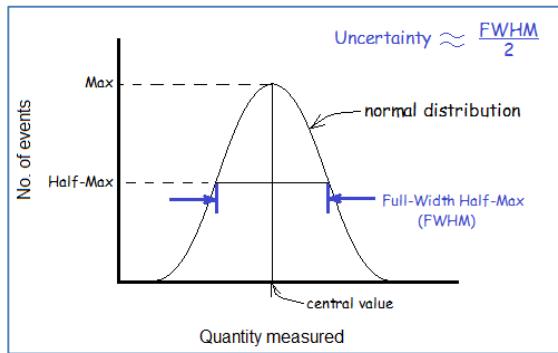


Figure 3: Defining uncertainty for a Gaussian distribution.

In conclusion, the uncertainty for each bin is found by using the standard deviation Poisson distribution \sqrt{N} . The uncertainty of the histogram is found using the approximation of the standard deviation of the Gaussian distribution FWHM/2.

In the student pages, your students have a histogram for class data from the *Rolling with Rutherford* activity in which marbles are rolled to try to knock hidden marbles out of the way. Each student makes 10 tries and records the number of hits. Figure 4 shows that histogram.

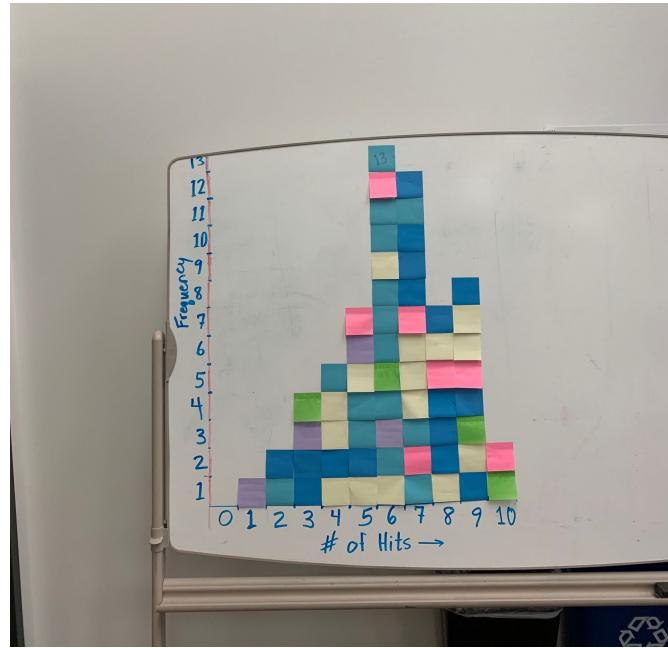


Figure 4: Histogram of Rolling with Rutherford data.

Figure 5 shows the histogram with bin error bars.

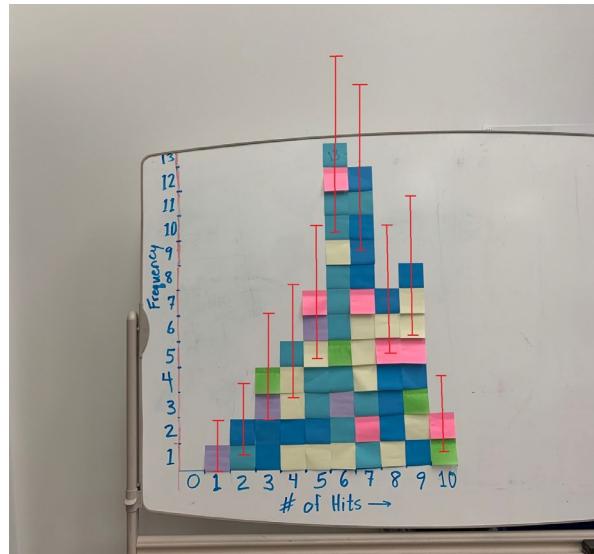


Figure 5: Histogram of Rolling with Rutherford data with bin error bars.

Figure 6 shows the histogram with FWHM identified.

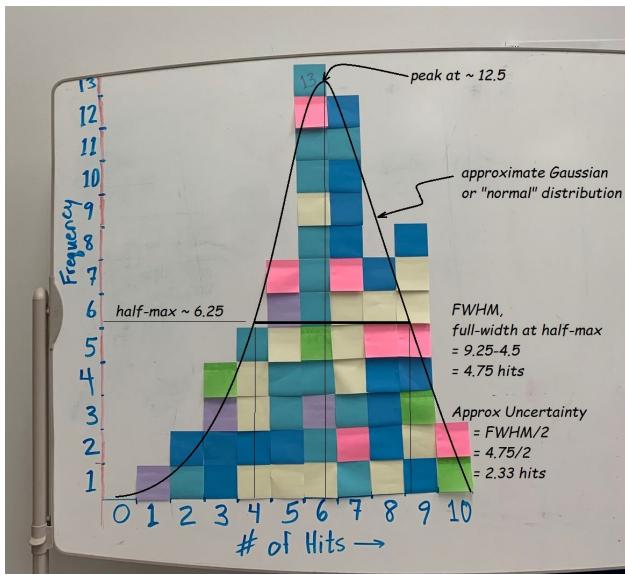


Figure 6: Histogram of Rolling with Rutherford data with bin error bars.

ASSESSMENT

You can use group work followed by classroom discussions for formative assessment. A sample summative assessment follows. The histograms for analysis are on the last page of the document.

Student Instructions:

For Histogram 1:

- **Calculate** the uncertainty of each bin.
- **Draw** the error bars on each bin.

For Histograms 2 and 3:

- **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 particle mass in GeV including uncertainty. Support your claim with evidence and reasoning.

Answer Key:

- **Determine** the uncertainty of the peak using FWHM.

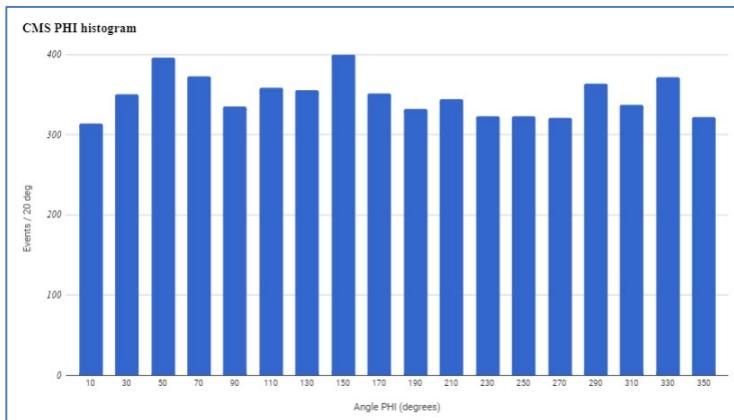
Histogram 2:

- The peak mass value is approximately 90 GeV at a frequency of 52. Half of 52 is 26. The values for frequency of 26 are 85 GeV and 95 GeV. Therefore, the width of the peak is 10 GeV. Half of that value is 5 GeV. Therefore, the best answer for the invariant mass at the peak is $90 \text{ GeV} \pm 5 \text{ GeV}$.

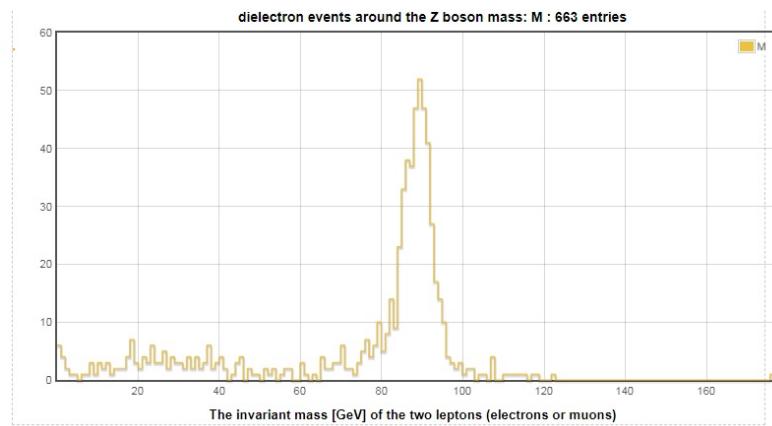
Histogram 3:

- The peak is approximately 42 GeV at a frequency of 52. Half of 52 is 26. The values for frequency of 26 are 30 GeV and 45 GeV. Therefore, the width of the peak is 15 GeV. Half of that value is 7.5 GeV. Therefore, the best answer for the transverse momentum at the peak is $42 \text{ GeV} \pm 7.5 \text{ GeV}$.

Histogram 1: Angle measurements from the CMS detector



Histogram 2: Mass plot from CMS



Histogram 3: Mass plot from CMS

