In this lecture we learn what physicists do with the 0.001% of the data that we do keep from LHC collisions. The simplest way to explain our process is with the scientific method, which you learned in elementary school. We ask a question: How do particles get mass? Posit a hypothesis: Through the Higgs field. Develop a test: If the Higgs field exists then I should find pairs of Z bosons with the invariant mass of the Higgs. Do the test: Turn on the LHC. Analyze the data. And that last step is where we’ll spend all of our time today.

Analyzing the data is a complex process with many steps. But, at its most basic level, we compare the properties of many events with expectations we derive from the Standard Model, and determine if we see what we expect or if we see something new. To be a bit more specific, let’s look at the case of the Higgs boson. We choose a particular property of an event, say the invariant mass of two Z bosons, record the value of this property is for very many events, and try to determine if statistically the set of invariant masses we measure is what we would expect from the Standard model with a Higgs boson or without a Higgs boson. Usually, we can’t tell if any particular event contains a Higgs Boson, but if we look at many of them, we can tell if some fraction of them seem to contain a Higgs.

Our expectations are determined using Monte Carlo simulations. These simulations have a few parts. First, we have the theory input. Our theorist friends have very detailed calculations of the properties of events should be for both the

http://hep.uchicago.edu/~tompkins/
Standard Model and for their favorite new theories. They tell us how often two quarks produce Z bosons in the standard model, and they tell us how often the invariant mass of the Z bosons is 120 GeV, 121 GeV, 122 GeV, etc.... We then randomly sample their prediction by rolling an “invariant mass” dice which is weighted such that 120 GeV will come up with the frequency they told us, 121 GeV will come up with its frequency, etc. Then we take that prediction and run it through a simulation of the ATLAS detector to see how we would measure those events. This process gives us our predicted distribution of events (like the blue in the Figure on the first page).

Sometime it is too difficult to simulate enough events, or simulate the events properly, so we also use the data itself for the predictions. In the case of the two photons, the Standard Model predicts a smoothly falling distribution of invariant masses, so we draw a line through those and see if anywhere the line looks like it has a bump in it. That bump is our Higgs signal!

When we see a signal, we then have to determine how confident we are that we are seeing something new. We do this by calculating the probability that the number of events we observed with a mass near 125 GeV could have come from our prediction without a Higgs boson. To say that we have “evidence” for a new particle, that probability has to be 0.3%. To say that we have “discovered” a new particle, as ATLAS and CMS said on July 4th, 2012, that probability has to be 0.00003% (less than 1 in a million chance!). The reason the criteria for discovery is so high is because we sometimes make mistakes and the background prediction isn’t gaussian distributed (i.e., follows the distribution in figure above) so in reality, the probability for the background to fluctuate and fake the signal is higher than 0.3%. But one in a million such a high threshold, that even if it isn’t really 1 in a million, but 1 in 500,000, thats still a high enough bar to say we have discovered something.

Some references on statistics:

http://blogs.scientificamerican.com/observations/2012/07/17/five-sigmawhats-that/

http://hep.uchicago.edu/~tompkins/