After flipping a coin 100 times, you tally up 42 heads and 58 tails. You expected a 50/50 distribution. Does this mean the coin is unfair? This video is part of the Probability and Statistics video series. Many events and phenomena are probabilistic. Engineers, designers, and architects often use probability distributions to predict system behavior. Hi, my name is Lourdes Aleman, and I am a research scientist in the HHMI Education Group in the biology department at MIT. I also work in the MIT Office of Educational Innovation and Technology. I helped develop the Star Genetics software featured in this video. Before watching this video, you should know what a probability distribution is and be familiar with basic genetics vocabulary including genotype, phenotype, homozygous, and heterozygous. You should also know how to use Punnett squares to predict the expected results of Mendelian crosses. After watching this video, you will be able to apply Chi square hypothesis testing to experimental data obtained from genetic experiments. Going back to question of whether or not our coin is fair, it turns out there is a statistical tool that can help us. Statistical tests can be used to determine if sample data can be assumed to come from a population that has a certain distribution. Today, we will use a statistical test called the Chi Square Test for Goodness of Fit. The Chi Square test is used to analyze categorical data. In other words, the test compares a sample of data collected about an event to expected values. In our coin-flipping example, our categories are heads and tails. An event is one coin flip. Our sample is 100 events. We expect that the coin should have a 50/50 chance of being heads or tails in any given event. The question we are trying to answer is – does our sample come from the 50/50 population distribution we expect? What we have just done is state our null hypothesis about the coin flipping experiment. The Chi square statistic tests this null hypothesis. We also need to state an alternative hypothesis. The alternative hypothesis is simply that the data do not come from the specified 50/50 population distribution. Before we can test our hypothesis, we need to calculate a Chi square statistic. This statistic measures the discrepancy between observed counts and expected counts. Each term in our summation finds the deviation, or error, between observed and expected values for each category. This deviation is squared to obtain positive values. Otherwise, everything would sum to zero because both our observed and expected outcomes sum to the same value. Each term in the summation is divided by the expected value of the outcome as a way of normalizing each difference. The Chi square statistic lives in a distribution. This is the Chi square distribution for any sample that has two categories of data, or, in other words, one degree of freedom. If the outcome of an event can only fall into one of two categories, we only need to know how many events fell into one category in order to determine how many fell into the other category. The Chi square distribution is actually a set of distributions, each for a different number of degrees of freedom. If we look at these distributions, the x-axis is chi-squared values and the y-axis is relative frequency. You might also notice that these distributions are positively skewed. This makes sense for a couple of reasons: The Chi squared statistic is always positive. Larger and larger Chi squared values indicate larger discrepancies between observed and expected values. While possible, these large discrepancies are less probable if the null hypothesis is true. If we calculate the Chi square test statistic for the coin flipping experiment, we get 2.56. How do we interpret this number? First, we need to look at the appropriate Chi-squared distribution. In the coin flipping experiment, there are two possible outcomes: heads or tails. If we know how many heads were flipped, we can calculate the number of tails flipped. So, we only have one degree of freedom. Let's refer to the Chi square distributions for one degree of freedom. We can compare our test statistic to the distribution to see if we have support for our null hypothesis or if we need to reject it. What will our criterion for support or rejection of our null hypothesis be? While this is a fairly arbitrary choice, in many fields, a result is said to not differ significantly from expectations if it has a 1 in 20 chance of happening. In other words, if the difference between the observed result and the expected result is small enough, we would expect to see it 1 in 20 times, which is a probability of 0.05. This probability is called the level of significance. If a Chi square value has a probability of occurring that is greater than our level of significance, it lends support to our null hypothesis. If a Chi square value has a probability of occurring that is less than our level of significance, we should reject the null hypothesis. Using the Chi square distribution for 1 degree of freedom, a chi square value of 2.56 has a greater than the 5% probability of occurring that we had set as our minimum. This means that we fail to reject our null hypothesis. In other words, what we observe is close enough to what we expect that we have confidence that our coin is fair. Now, suppose you are planning an experiment where you will cross mutant flies that are heterozygous for a wingless phenotype to each other. The wingless phenotype is dominant relative to the wild-type winged phenotype. Let's use the notation Gg for the genotype of the parent mutant flies. In this case, the wild type, winged flies, are homozygous recessive. What percentage of flies resulting from a cross of two heterozygous wingless mutants would you expect to be wild type versus mutant? Pause the video here and use a Punnett square to justify your answer. Performing this cross using a Punnett square, you would expect a 3:1 ratio of mutant to wild-type flies. You perform the cross with the parental mutant flies. The cross produces 50 progeny. 32 are mutant and 18 are wild type. So, only 64% of our flies are mutant, even though you expected 75% of them to be mutant. Although it is not what you expected, it doesn't seem that far off either. You decide to continue mating your parental mutant flies until you have 1000 progeny. Surely, with this large a sample size, the ratio of mutant to wild type flies will be closer to what you expect. Voila. Oh, wait a minute. Even with a 1000 progeny, our experiments still produce 68% mutant and 32% wildtype flies. Now you are really starting to wonder if you can call this result "close enough" to the expected outcome. We can use the Chi-square test to evaluate the hypothesis that the observed ratio of 68:32 is the same as the expected ratio of 3:1. Pause the video here, calculate the Chi square test statistic, and determine the degrees of freedom in your data. You should have obtained 29.2 for your Chi square value. Because there are only two phenotypic outcomes for your cross, there is only one degree of freedom. Any fly that isn't a mutant is wild type and vice versa. The Chi square distribution can also be presented in tabular form. If you look at this table, our experiment with one degree of freedom and a Chi square value of 29.2 corresponds to a probability that is much smaller than 0.05. Do you have support for the

Chi square value of 2512 corresponds to a probability that is much smaller than 0.05. Do you have support for the null hypothesis or should you reject it? Pause the video here and take a moment to think about it. Because the ratio of progeny that you obtained in your experiment has a probability of happening that is less than our level of significance, you should reject our null hypothesis. This indicates that our discrepant results are due to something more than just random chance. A multitude of other factors could be contributing to our results. For example, there may be interactions with other genes that you are not aware of. Or, environmental factors could be playing a role. All we know is that our null hypothesis of a 3:1 ratio of mutant to wildtype flies is incorrect. As a next step, you decide to go back and review your notes to see if you can formulate another plausible hypothesis. Reviewing your Punnett square, you expected to obtain this genotypic ratio in the F1 progeny. If you were to select a GG mutant from the F1 progeny and cross it with a gg wild-type fly, you would expect that all of the progeny would be mutants. After attempting this cross with a variety of flies from the F1 progeny that have a mutant phenotype, it becomes clear that you cannot find a mutant fly that when crossed with a wild-type fly, produces all mutant progeny. Somehow, you keep selecting heterozygous mutants. You begin to suspect that the F1 population does not contain the expected number of GG flies! Checking your lab notebook, you see a note about an observation that you barely paid attention to at the time. When you mated the two parental mutant flies, there were a number of dead embryos in your vial. A new hypothesis is forming in your mind. Can you come up with a hypothesis that may explain the observations you have made through these experiments? Pause the video here. Discuss and justify your hypothesis with a classmate. While there are many possible hypotheses, one possible hypothesis is that the dead embryos in your vial were the missing GG mutants. It may be that the wingless trait when homozygous, results in lethality. This would result in a phenotypic ratio of 2:1 mutant to wild-type flies in the F1 progeny. Again, you can use the Chi square test for goodness of fit to test your new null hypothesis. Going back to your data of 1000 progeny, use the test to evaluate the null hypothesis. Pause the video. You should have obtained a Chi square value of 0.36. For one degree of freedom, you fail to reject your null hypothesis. While you have support for your null hypothesis, more experiments are needed to confirm the idea that the GG genotype is lethal as you suspect. The results from our Chi squared test have led us to a reasonable next experiment to perform. In this video, you saw how the Chi-square statistic can help us determine if our experimental genetic data came from the expected population distribution. Comparing observed values to expected values helped us determine if deviations in our data could be explained by random chance or if our hypothesis needed revision. The Chi-Square Test for Goodness of Fit is just one statistical test of many that allows us to determine how confident we are that a sample of data comes from a population with a certain distribution. It is important to understand what these tests mean and how to interpret their results, so that you know their limitations and can make informed decisions.