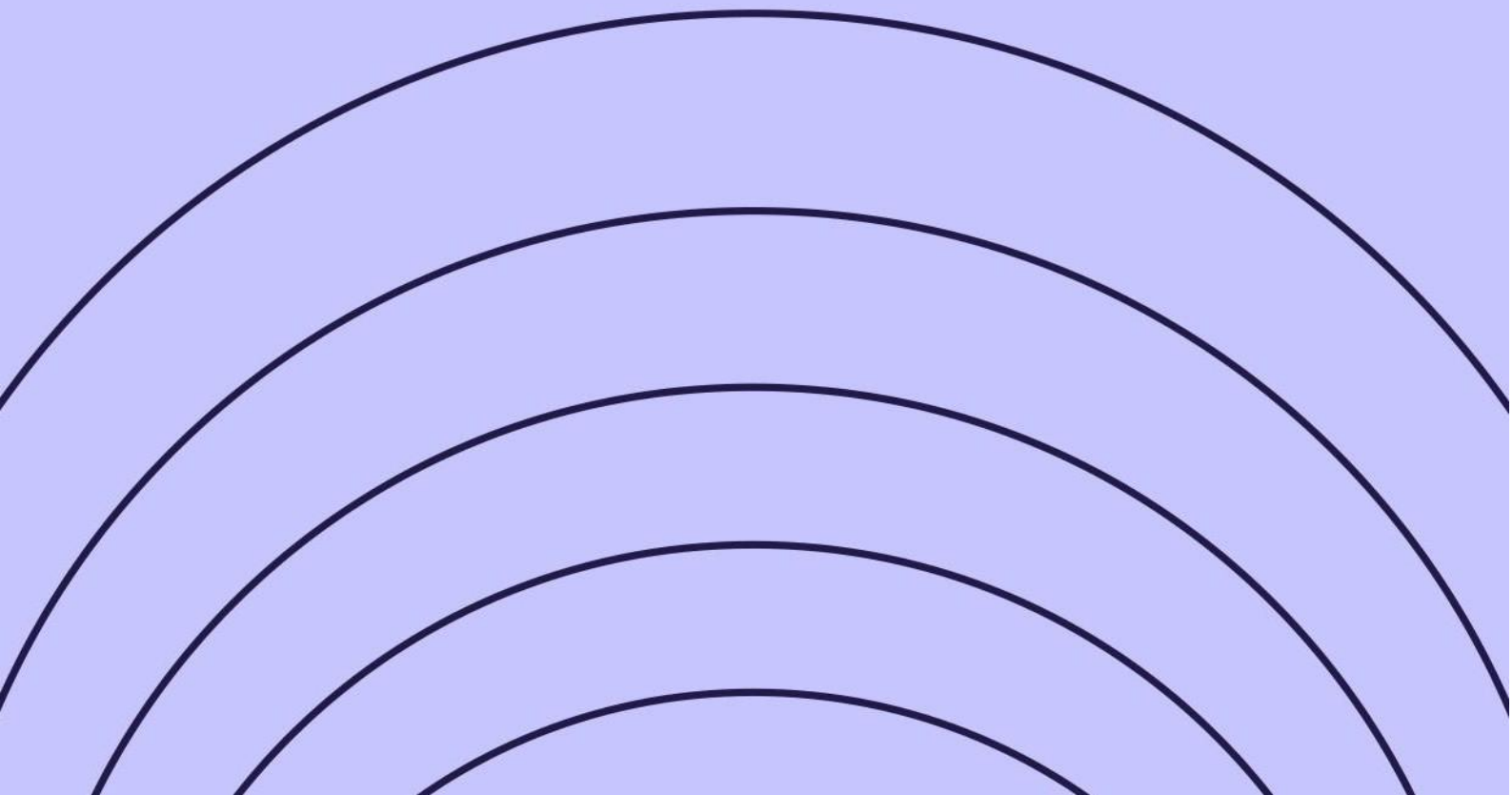
Pearson

Video transcript

Chi-Squared Test



Transcript

In A-level biology there are various statistical tests that you must be aware of and know when to use them correctly. In this video, I will focus on the chi-squared statistical test.

On the screen you can see the equation for chi squared. We will look at this in more detail later in the video.

Use the chi-squared test when your data is categoric, like counting phenotypes in a genetic cross. The chi-squared test is used to compare the observed frequencies from an experiment with the expected frequencies predicted by a theory.

You should be able to calculate these expected values either by assuming an equal distribution across categories, or by applying a theoretical model such as Mendel's law of inheritance.

The chi-squared test evaluates the null hypothesis, which states that there is no significant difference between the observed and expected frequencies.

When conducting any statistical test, there are several key steps to follow. The first is to state the null hypothesis. A null hypothesis states that there is no correlation between two variables. You then conduct the statistical test to either accept or reject the null hypothesis.

The second step is to carry out the chi-squared equation. Using the equation, you calculate the value for chi squared.

Step three is to identify the critical value where p is equal to 0.05. Using critical value tables that will be provided in your exam, you find the critical value at a significance level of 0.05. This is the normal value used in biology.

Step number four is then to compare the values. You then compare the value for chi-squared with the value for p to decide if it is greater or less than p.

The fifth stage is to generate a conclusion. This will then allow you to accept or reject the null hypothesis and write a conclusion.

Now let's look at a past paper question.

In this question. You're told that cauliflowers can be light purple, dark purple, or white. You're told that the colour in cauliflowers is controlled by a single gene with two alleles. The purple colour has the allele capital P small r. The white colour has the allele small p small r. You're told that in an investigation dark purple cauliflowers were crossed with purebred white cauliflowers, and that the F1 generation were all light purple. You were then asked to state the genotype of the F1 generation.

In this question, you're required to carry out a genetic cross with Pr, Pr and small pr, small pr. Like so. So therefore, you end up with all of the genotypes being the same where you have capital P small r small p small r. So, they all end up with the same genotype.

The next part of the question requires you to work out the expected number of the three phenotypes when the F2 generation self-fertilize. This requires you to complete another genetic cross. As you can see, I have already completed the genetic cross and from completing this genetic cross, we can identify that there are three different genotypes: capital P small r capital P small r, and there is one; capital P small r small p small r, of which there are two; and small p small r small p small r, of which there is one. Therefore, the actual ratio is 1 to 2 to 1.

Given that there is a total of 103 cauliflowers observed and based upon the 1 to 2 to 1 ratio, the expected numbers are: capital P small r capital P small r, which is dark purple, would be 26. Capital P small r small p small r, which is light purple, is 52. And small p small r small p small r, which is white, there are also 26. Therefore, the ratio, the expected ratio, is 26 to 52 to 26. So we can see from the mark scheme that we have identified the number of each calculated - the 1 to 2 to 1 ratio - and the number of each phenotype identified as 26 dark purple, 52 light purple and 26 white.

In this part of the question, we are asked to calculate the value for chi-squared using the formula.

Now let us have a look at the equation a little bit more detail. The first symbol that we see here is the symbol for chi-squared itself. The second symbol, sigma, refers to the sum of. The letter e is referring to the expected number, and the letter O is referred to as the observed number.

The easiest way to calculate the chi-square is to complete it using the table on screen. I have already put in the data for you, but if we look specifically just at the first phenotype, which is capital P small r capital P small r. We have already filled in the observed number from the question as 20, which is O. We have already filled in the letter E, which is the expected number we calculated from the previous question, 26.

The first calculation is O minus E. So, in this case 20 -26 which gives us a value of minus six. As you can see in the next stage, we are simply squaring this. So, it is minus six squared which is 36.

And then in the last part you can see we are then dividing this answer by E. So therefore, that is 36 divided by 26. This gives us a value of 1.385.

Once all the data is inputted and completed in the table, the final thing to do is to add together the figures in the final column, and this gives us a value of 2.366 rounded up to 2.37. This is the value for chi squared.

As you can see from this question, from the mark scheme, there are three marks available. The first mark is for completing the calculation of the observed minus expected. And you can see this is the three values for each phenotype.

The second mark is for squaring those values. And then dividing by E and you can see then the three separate answers for each of those of the phenotypes.

And then the final mark is for your chi-squared value calculated correctly.

From the mark scheme, it's important to note that you can gain full marks with the correct answer for chi-squared on its own. However, if you make a mistake and provide the incorrect answer if you've shown the steps that you've used, you may still be able to gain some marks.

The final part of the question is requiring you to do three things. The first thing is to find the critical value and compare it to chi squared. The second thing is to either accept or reject the null hypothesis. And the third thing is to explain why the conclusion provided is likely to be valid.

The value you calculate for chi-squared was 2.37. We need to compare this to a critical value from the table provided in the question. To identify the critical value from the table, we need to understand the degrees of freedom column first.

The degrees of freedom are the number of possibilities minus one. We often refer to this as n minus one.

In this scenario there were three possible phenotypes. Therefore, the degrees of freedom are three minus one, which equals two.

In biology we always use p is equal to 0.05 as the level of significance. Therefore, the critical value for our results is 5.991.

If the chi-squared is greater than critical value, we reject the null hypothesis. If the chi-squared is less than the critical value, we accept the null hypothesis. In this case, we would accept the null hypothesis because the value for chi-squared is less than the critical value.

Therefore, the conclusion provided is likely to be valid, and the differences between the observed and expected are due to chance.

As you can see from the mark scheme, there are three marks available for this question. The first mark is for identifying the critical value at P equal to 0.05, and the value is 5.991. The second mark is for identifying that the calculated value is less than the critical value, so that your value for chi-squared is less than the critical value. And your final mark, your third mark, is for identifying that: therefore, the conclusion is likely to be valid.

Transcript

© Pearson Education 2025