

# Chi-Square

How do you know if your data is the result of random chance or environmental factors?

## Why?

Biologists and other scientists use relationships they have discovered in the lab to predict events that might happen under more real-life circumstances. At some point those predictions are either supported or not by real data. It would be nice if the two sets of data matched exactly, but more often than not they don't. This could be attributed to random fluctuations in the variables, or it could be due to a factor that the scientists overlooked. How does the scientist know? There are many statistical calculations that help answer this question. One that is commonly used by biologists is **Chi-Square**.

## Model 1 – Calculating Chi-Square ( $\chi^2$ )

Hypothesis: There is an equal chance of flipping heads or tails on a coin.						
<b>Coin A</b>						
	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$	$\chi^2 = \sum \frac{(o - e)^2}{e}$
Heads						
Tails						
Total	200	200				
<b>Coin B</b>						
	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$	$\chi^2 = \sum \frac{(o - e)^2}{e}$
Heads						
Tails						
Total	200	200				

1. What is the hypothesis that has been tested in Model 1?
2. Describe in one or two complete sentences the experiment being performed in Model 1.
3. How many flips of the coin will be conducted in each trial for the experiment in Model 1?

4. How many of the 200 flips would you expect to be heads and how many would you expect to be tails? Fill in the expected (e) column of Model 1 for both coin A and coin B.
5. Assume the experiment for coin A resulted in 108 heads and 92 tails. Fill in the data table for coin A under observed (o).
6. The experiment was repeated with a different coin. coin B resulted in 120 heads and 80 tails. Fill in the data table for coin B under observed (o).
7. If you were told that one of the coins used in the experiments in Model 1 was a “trick” coin, which coin would you predict was rigged? Explain your reasoning.

## Read This!

---

The experiments in Model 1 did not give the expected outcome, so the question becomes: Is this due to random chance, or does the coin being flipped favor heads for some reason? The statistical calculation of **chi-square** ( $\chi^2$ ) will help determine if there is a significant difference between the expected outcome and the observed outcome. In statistics a “significant” difference means there is less than 5% chance that the variation in the data is due to random events. Therefore, the variation is most likely due to an environmental factor.

---

8. What symbol represents chi-square?



9. Use the equations in Model 1 to complete the calculations of  $\chi^2$  for coin A and coin B.



10. Circle the correct phrase to complete the sentence.  
A larger chi-square value means the observed data is (very different from/very similar to) the expected data.
11. One hundred heterozygous (Bb) males mate with one hundred heterozygous (Bb) females.
  - a. Draw a Punnett square to show the possible genotypes of the offspring from each pairing.
  
  
  
  
  
  
  
  
  
  
  - b. Predict the number of offspring from the 100 mating pairs that will be each genotype.  
BB =                      Bb =                      bb =

c. The observed outcome from this experiment is 28 BB offspring, 56 Bb offspring and 16 bb offspring. Use a table similar to that in Model 1 to calculate chi-square for this experiment.

d. In your opinion, is the observed data significantly different from the expected data in this mating experiment? Justify your answer.

## Read This!

---

To determine if the chi-square value is large enough to be “significant,” researchers need one more thing—**degrees of freedom**. If an experiment has five possible equivalent outcomes, then in reality there is one result and four additional possible results, making a total of five. The degrees of freedom are always calculated as the number of possible outcomes minus one.

---

12. Consider the coin flip experiments in Model 1 where you could get heads or tails.

a. How many outcomes were possible in the coin flip experiments?

b. How many degrees of freedom were there in the coin flip experiments?

13. Consider the mating organisms in Question 11.

a. How many genotype outcomes were possible in that experiment?

a. How many degrees of freedom were there in that experiment?

## Model 2 – $\chi^2$ Analysis

Degrees of Freedom	$\chi^2$ Values										
	1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
<b>P Value =</b>	<b>0.95</b>	<b>0.9</b>	<b>0.8</b>	<b>0.7</b>	<b>0.5</b>	<b>0.3</b>	<b>0.2</b>	<b>0.1</b>	<b>0.05</b>	<b>0.01</b>	<b>0.001</b>
<b>% probability</b>	<b>Not Significant</b>							<b>Significant</b>			

14. The table in Model 2 has three types of values— $\chi^2$ , degrees of freedom, and P values or probabilities.
- Which of these are along the bottom of the table?
  - Which of these are along the side of the table?
  - What chi-square value is needed to have a P value of 0.5 in an experiment with two degrees of freedom?
15. The table in Model 2 is a reference table used by scientists to interpret the calculated chi-square value for their experiment. It converts the chi-square value into a probability that the differences in the data are only due to chance.
- Which P Values in the table indicate that the difference between the expected and observed data was likely due to chance?
  - Which P Values in the table indicate that the difference between the expected and observed data was likely due to an environmental factor?
16. What does a P value of 0.7 mean in terms of percent chance that the data sets are different only by chance?

17. As the chi-square values get larger, is the data more likely to be significant or not significant?
18. When an experiment has more degrees of freedom, is a larger chi-square needed for a significant outcome, or is a smaller chi-square needed for a significant outcome?
19. Which row of  $\chi^2$  values in Model 2 should be used for the coin flip experiments?

20. Find the P values for each coin experiment in Model 1 using the  $\chi^2$  values in Model 2.

*Note:* If the  $\chi^2$  value falls between two points on the chart, give a range of P values.

a. Coin A =

b. Coin B =



21. Consider your answers in Question 20,

a. Which coin produced a nonsignificant  $\chi^2$  value, indicating that the outcome was not statistically different from what was expected?

b. Which coin produced a significant  $\chi^2$  value, indicating that the outcome was statistically different from what was expected and therefore due to an environmental factor?

c. With your group propose an explanation for why the results of either coin were statistically significant.

22. Does the mating experiment from Question 11 indicate that an environmental factor was affecting the outcome? Justify your answer using information from Model 2.

23. A researcher is investigating a flowering plant. The purple flower allele,  $P$ , is dominant to the white flower allele,  $p$ . A cross was performed between a purple-flowered plant and a white-flowered plant. The 206 seeds that were produced from the cross matured into 124 purple-flowered plants and 82 white-flowered plants. Construct a table similar to those in Model 1 to calculate chi-square for the null hypothesis that the purple-flowered parent plant was heterozygous.



## Extension Questions

24. You have performed a dihybrid cross of plants and got the following data: 206 purple tall, 65 white tall, 83 purple short, 30 white short.

a. Using your knowledge of genetics, determine the expected outcome for this experiment.

b. What are the degrees of freedom in this experiment?

c. Calculate chi-square for your experimental data below.

	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$
Purple and tall					
Purple and short					
White and tall					
White and short					

d. Does your collected data differ significantly from the expected values? Explain.