

# Key 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.						
Coin A						
	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$	$\chi^2 = \frac{\sum(o - e)^2}{e}$
Heads	108	100	8	64	0.64	
Tails	92	100	-8	64	0.64	
Total	200	200				1.28

  

Coin B						
	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$	$\chi^2 = \frac{\sum(o - e)^2}{e}$
Heads	120	100	20	400	4	
Tails	80	100	-20	400	4	
Total	200	200				8

1. What is the hypothesis that is being tested in Model 1?  
*There is an equal chance of flipping heads or tails with a coin.*
2. Describe in one or two complete sentences the experiment being performed in Model 1.  
*A coin is being flipped, and the number of times it lands on heads and tails is recorded.*
3. How many flips of the coin will be conducted in each trial for the experiment in Model 1?  
*200*

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.

*See Model 1.*

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).

*See Model 1.*

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).

*See Model 1.*

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.

*Coin B had a result that was further from the predicted result, so it is likely that it is the rigged coin.*

## 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?

$\chi^2$



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

*See Model 1.*



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	b
B	BB	Bb
b	Bb	bb

- b. Predict the number of offspring from the 100 mating pairs that will be each genotype.

BB = 25      Bb = 50      bb = 25

- 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.

	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$
BB	28	25	3	9	0.36
Bb	56	50	6	36	0.72
bb	16	25	-9	81	3.24

*Chi-square* = 4.32

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

*Answers may vary. Some students will say that the observed data varies from the expected data enough that it cannot be due to chance. Others will say they are close enough to be random. The point to be made here is that they are just making a judgment call and they need the statistical analysis to know for sure.*

## 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.
- How many outcomes were possible in the coin flip experiments?  
*Two.*
  - How many degrees of freedom were there in the coin flip experiments?  
*One.*
13. Consider the mating organisms in Question 11.
- How many genotype outcomes were possible in that experiment?  
*Three.*
  - How many degrees of freedom were there in that experiment?  
*Two.*

## 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	10.83	
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	
P Value =	0.95	0.9	0.8	0.7	0.5	0.3	0.2	0.1	0.05	0.01	0.001	
% probability	Not Significant								Significant			

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?  
*P values or probabilities.*
  - Which of these are along the side of the table?  
*Degrees of freedom.*
  - What chi-square value is needed to have a P value of 0.5 in an experiment with two degrees of freedom?  
*1.39*
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?  
*P values from 0.95 to 0.1 are considered “not significant” and therefore only 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?  
*P values less than 0.05 are considered “significant” meaning the data is 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?  
*It means there is a 70% chance that the variability in the data is only due to chance. This is not significant.*

17. As the chi-square values get larger, is the data more likely to be significant or not significant?  
*As the chi-square values get larger, the data is more likely to be significantly different from the expected values.*
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?  
*With more degrees of freedom, the chi-square value must be larger to achieve the same P value.*
19. Which row of  $\chi^2$  values in Model 2 should be used for the coin flip experiments?  
*The top row with one degree of freedom.*
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.*
- Coin A = *between 0.2 and 0.3.*
  - Coin B = *between 0.01 and 0.001.*
21. Consider your answers in Question 20.
- Which coin produced a nonsignificant  $\chi^2$  value, indicating that the outcome was not statistically different from what was expected?  
*Coin A.*
  - 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?  
*Coin B.*
  - With your group propose an explanation for why the results of either coin were statistically significant.  
*Accept any reasonable answer such as the coin was weighted unevenly causing it to flip and show heads more often.*
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.  
*Chi-square of 4.32 with two degrees of freedom gives a P value between 0.2 and 0.1—this is not significant.*

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.

	P	p
P	Pp	pP
p	pP	pp

	Observed data (o)	Expected (e)	(o - e)	(o - e) <sup>2</sup>	$\frac{(o - e)^2}{e}$
Pp	124	103	21	441	4.28
pp	82	103	-21	441	4.28

Chi-square = 8.56

*This chi-square with one degree of freedom gives a P value of 0.01 to 0.001, so it is significant. Therefore, the hypothesis that the purple-flowered parent was heterozygous is false. There must be some other factor affecting the outcome.*



## 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.

216 purple/tall      72 purple/short      72 white/tall      24 white/short

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

Three.

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	206	216	-10	100	0.463
Purple and short	83	72	11	121	1.681
White and tall	65	72	-7	49	0.681
White and short	30	24	6	36	1.5

Chi-square = 4.325

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

*P is between 0.2 and 0.3 so the difference between observed and expected is not significant. Thus, the experiment yielded the expected results (i.e. the observed data matches the expected values under the rules of Mendelian genetics).*