This is a simple test to determine whether to events are correlated. In this case, whether the condition to which you exposed the flies affects the distribution.

first you calculate the χ^2 Value, then compare it to a table of values to determine if there is a correlation.

The general equation is:

$$\chi^2 = \sum \frac{(O-e)^2}{e}$$ that is: sum of $\frac{(\text{observed}-\text{expected})^2}{\text{expected}}$. one term for each measurement.

Let's take a simple case: I put some ethanol on a cotton ball on one side of the chamber. My hypothesis is: flies will be attracted to low concentrations of ethanol (rotting- or, fermenting-fruit results in ethanol production).

I have 36 flies in the chamber. I expect 18:18. But, 19:17 or even 20:16 would not be surprising. Would 21:15 prove the flies have a preference? Or, would I need to see an even bigger difference to be convinced?

That's where the χ^2 test comes in.

It allows you to determine how likely the distribution you observe would be if there were no preference.

a Coin flip.

If I flip a coin 10 times, what are the chances of getting 6 heads and 4 tails. We can actually put a number on that. It's about 0.65. It's actually slightly <u>more</u> likely than 5:5. Note that it's not a matter of percentages. The probability of 60 Head 40 Tail is much lower (about 0.15).

This is an important point: for a chi-square (χ^2) test, the actual count, not % is used. Therefore, <u>the larger your data set, the more confident you can be of your result.</u>

Sample Calculation:

Let's suppose in your experiment, you saw 22 flies on one side an 14 on the other.

$$\chi^2 = \frac{(22-18)^2}{18} + \frac{(14-18)^2}{18} = 1.78$$

Some of you may be surprised that I included 2 terms, one for each side. After all, if there are 22 flies on one side, I don't have to measure the other side. Does it really count as a separate measurement? Yes it does... but, you are right in that it is not an "independent" measurement. We will account for that in the next step by setting something called the "degrees of freedom." If, as in this case, there are only two classes of

outcome (on one side or the other), once I count one side, I know the other. We say there is only one degree of freedom (DF). I there were 3 classes (say, a third section of the chamber), as soon as I measured the 1st two, I would know the third. So, that would be 2 DF. However many classes of result there are, subtract one to get the DF. For this experiment, DF will be 1. later in the year when we do genetics, there will typically be 4 classes of result, so 3 DF.

The Null Hypothesis: (this is important)

We saw a distribution in our example of 22:14.
We think we caused it using some condition. The null hypothesis is that <u>we are wrong</u> - it was just a random variation. To determine if that is likely, we compare our χ^2 value of 1.78 to a table of values.

for 1 DF, our χ^2 value has a probability between 0.1 and 0.25 of ocurring randomly (1.78 falls between 1.32 and 2.71. I used a more complete online tool

Accept Hypothesis ← → Reject Hypothesis

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09

Our Value

Chi-square table.

https://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html

to determine our probability at 0.18.

This means there is a probability of 0.18 that our 22:14 distribution was random.
That number does indicate it is not likely to be random.
However, the result is not definititive enough to <u>reject</u> the null hypothesis. It may be true that flies prefer ethanol - but I can't be certain. We like to see p-values < 0.1 or even 0.05 to reject the null Hypothesis.

It is not a linear function. What if our distribution was 24:12? That's just 2 more flies on the experimental side. Calculate X^2 and the probability of the null hypothesis.

We will discuss in class examples of how to include more data.