**The Chi-Squared (*χ2*) Test**

The *χ2* test is a statistical test to see if *observed* values from a data set are similar to or significantly different from *expected* values. This test is used when observed data fall into categories or classes (ex. the number of students wearing flip flops vs. sneakers) not continuous measurements (ex: weight or length, such measurements are analyzed using different statistical procedures like t-tests). You are using the *χ2* test to determine the probability that your data set (observed) was the result of random chance (expected). The expected data values are based on a mathematical model, and are used to create the null hypothesis. After the value of *χ2* is calculated, a *χ2* table is used to find the *χ2* critical value.

* If the calculated is greater than the critical value, we say that there is a significant difference between the observed and expected values.
* If the calculated is less than the critical value, we say that there is no evidence supporting a significant difference between the observed and expected values.
* Finding the critical value requires knowing the degrees of freedom and the significance level
  + Degrees of freedom (*df*) – the number of pieces of information that are free to vary. For this module we will be using *df* = 1 (please **see note**)
  + A typical significance level is 5% = 0.05. This means we are 95% confident in our conclusion.
* Each calculated value is associated with a p-value which can also be directly compared to the significance level to make conclusions to the test (please **see note**)

**Calculating the *χ2*-Value**

* Record all observed values (O).
* List the expected values (E). Expected values are calculated by multiplying the probability of an outcome by the total number of outcomes.
* List the difference in the observed and expected values (O – E).
* List the values for (O – E)2.
* List the values for .
* Chi-Squared (*χ2*) is the sum of all the values of , where the Greek letter sigma ∑ represents “sum.”
* Compare the calculated and critical value

Example: M&Ms

There are 10 M&Ms in a container, 7 are blue and 3 are orange. We randomly selected two M&Ms at a time, using both hands (one M&M each), and recorded the results in the column labeled “Observed”. After each draw, we put the M&Ms back in the container. We did this 20 times. We drew two blue M&Ms (BB) 10 times, one blue and one orange M&M (OB or BO) 8 times and two orange M&Ms 2 times.

Was the combination of M&M’s we randomly drew similar to what we expected?

Our null hypothesis is randomness – we should get what is expected by the laws of probability. If we don’t, then something is going on (an alternate hypothesis).

Calculating Expected values: the probability of BB is (0.7)(0.7) = 0.49, OO is (0.3)(0.3) = 0.09, BO =(0.7)(0.3) = 0.21, OB =(0.7)(0.3) = 0.21, the probability of BO *or* OB (if we don’t care about order) is 0.21 + 0.21 = 0.42. Therefore, if you select 2 M&Ms 20 times, the **expected** value for getting BB is (0.49)(20) = 9.8. One would expect to draw BB 9.8 times.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Color of M&Ms** | **Observed (0)** | **Expected (E)** | **O – E** | **(O – E)2** | **(O – E)2/E** |
| BB | 10 | 9.8 | 0.2 | 0.04 | 0.004 |
| BO or OB | 8 | 8.4 | -0.4 | 0.16 | 0.019 |
| OO | 2 | 1.8 | 0.2 | 0.04 | 0.022 |
|  | | | | | ***χ2*** = 0.045 |

We will use *df* = 1 (**see note**). Go to the *χ2* critical value table at the bottom of the page and look up *df* = 1 and significance level of 0.05.

**Chi-Squared (*χ2*) Critical Values Distribution Table**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Significance Level** | | | | | | | |
| **Degrees of Freedom** | **0.99** | **0.80** | **0.50** | **0.20** | **0.10** | **0.05** | **0.01** | **0.001** |
| **1** | 0.00016 | 0.064 | 0.46 | 1.6 | 2.7 | 3.8 | 6.6 | 10.8 |
| **2** | 0.02 | 0.45 | 1.4 | 3.2 | 4.6 | 6.0 | 9.2 | 13.8 |
| **3** | 0.12 | 1.0 | 2.4 | 4.6 | 6.3 | 7.8 | 11.3 | 16.3 |
| **4** | 0.30 | 1.6 | 3.4 | 6.0 | 7.8 | 9.5 | 13.3 | 18.5 |

Evaluating the Hypothesis:

The critical value for *χ2* = 3.8. Since our *χ2* calculated value is 0.045 and less than 3.8, this means our p-value is much greater than 0.05, (in fact it is close to 0.8 – see the *χ2* critical value table. In other words, the probability that our observed results can be explained by randomness is approximately 80%).

Therefore, the combination of M&M’s we drew was very similar to what we expected. Which is what the null hypothesis stated.

We conclude that there is no evidence supporting a significant difference between the observed and the expected values. In other words, we fail to reject the null hypothesis.

What if? - What if there was a real difference between observed (what we drew) and what was expected? In order to conclude that the combination of M&M’s we drew was significantly different from expected (something was going on), we would need to reject no difference from expected (reject the null hypothesis). We can conclude this if our *χ2* calculated value is greater than the *χ2* critical value.

Summary:

Reject the null hypothesis – significant difference between the observed and the expected values (*p-value* < 0.05)

Fail to reject the null hypothesis – no significant difference between the observed and the expected values (*p-value* > 0.05)

**Practice Problem 1**

Put 8 blue and 2 green M&Ms in a container. Without looking, choose two M&Ms and record your results in a table as was done in the example above. Do this 20 times. Use Chi-Square to determine if there is a significant difference between the observed values and the expected values. Write your conclusion using complete sentences. (Degree of freedom = 1)

*Pieces of paper can be substituted for M&Ms – but M&Ms taste better.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Color of M&Ms** | **Observed (0)** | **Expected (E)** | **O – E** | **(O – E)2** | **(O – E)2/E** |
| BB |  |  |  |  |  |
| BG or GB |  |  |  |  |  |
| GG |  |  |  |  |  |
|  | | | | | ***χ2*** = |

Was the combination of M&M’s you drew similar to what was expected? In other words, did you draw them randomly?

Was there a significant difference between observed and expected?

Using the *χ2* critical value table – what is the approximate probability that your data set (observed) was the result of random chance (P-value)? Is p-value < 0.05?

The null hypothesis is that the combinations are drawn randomly and there is no significant difference between observed and expected. Do you reject or fail to reject the null hypothesis? Explain and justify your answer.

**Practice Problem 2**

Your imaginary younger sibling is going to help you this time. The problem is your sibling REALLY likes blue M&M’s (it is annoying). You put 8 blue and 2 green M&Ms in a container just like last time. You asked your imaginary sibling to choose two M&Ms randomly but you have your suspicions. The results are recorded in the table below. You drew M&M’s a total of 20 times. Use Chi-Square to determine if there is a significant difference between the observed values and the expected values. Write your conclusion using complete sentences. (Degree of freedom = 1)

*Pieces of paper can be substituted for M&Ms – but M&Ms taste better.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Color of M&Ms** | **Observed (0)** | **Expected (E)** | **O – E** | **(O – E)2** | **(O – E)2/E** |
| BB | 7 |  |  |  |  |
| BG or GB | 9 |  |  |  |  |
| GG | 4 |  |  |  |  |
|  | | | | | ***χ2*** = |

Was the combination of M&M’s you drew similar to what was expected? Did your sibling draw the M&M’s randomly or did they cheat and eat some blue M&M’s?

Was there a significant difference between observed and expected?

Using the *χ2* critical value table – what is the approximate probability that your data set (observed) was the result of random chance (P-value)? Is p-value < 0.05?

The null hypothesis is that the combinations are drawn randomly and there is no significant difference between observed and expected. Do you reject or fail to reject the null hypothesis? Explain and justify your answer.

**Note: Degrees of freedom (*df*)** – We are using a Chi-square ( *χ2*) test for a contingency table. (There is also a Chi-squared goodness of fit test, they are similar but the degrees of freedom are calculated differently.) Our contingency table has two categories; orange and blue. In a contingency table with r rows and c columns there are (r-1) x (c-1) degrees of freedom. Therefore in a 2X2 contingency table (2-1)X(2-1)= 1 *df*. Another way to think about it is because we have two categories our degree of freedom is 2-1 =1. We subtract 1 because if we know the relative frequency of one color M&M then we know the other. Therefore, only one color category that can vary, since the other is dependent on the first. One degree of freedom.

We draw two M&M’s at a time, using both hands.



P-value - Biological explanation for this scenario: P-Value is the probability that our observed values (our results) can be explained by the null hypothesis (randomness – consistent with laws of probability). Our null hypothesis is that the M&M’s are drawn randomly from the container and consistent with the laws of probability. That would mean we expect our observed frequencies to be as predicted by probability – the Expected values. If the p-value is really small, less than 0.05, that means that the probability our observed values can be explained by randomness is really, really small - so small that we can reject randomness as our explanation (null hypothesis). Therefore, we can accept our alternative hypothesis – something else is happening.

