Chi-square Analysis

Background: Chi-square analysis is a mechanism which statisticians and scientists use to determine whether the difference between our **expected** results and the results we **observe** are simply due to chance or have arisen for some other reason (because of a variable). For example, when rolling a die, we would expect that the chance of rolling 1, 2, 3, 4, 5, or 6 would be equal, 16.67% or 0.1667 when expressed as a frequency. However, if a weighted die or a flawed, unfair rolling technique is used, the observed data will not closely resemble the expected results. Using Chi-square analysis, we can statistically prove or disprove the existence of the aforementioned variables within a high degree of confidence (almost always 95% confidence).

Null hypothesis – a prediction that the introduced variable will have no effect, or that there will be no difference between the observed and expected values since the outcome was indeed due to chance alone

Ex.

> Question: When you roll a die 100 times, will the number of 1, 2, 3, 4, 5, and 6s that come face up be equal?
>
> Null hypothesis: There will be an equal number of 1, 2, 3, 4, 5, and 6s that come face up.

or

> Question: Does taking aspirin reduce your chance of having a heart attack?
>
> Null hypothesis: Taking aspirin will have no effect on the risk of heart attack.

In this situation, if the rate of heart attack between individuals who took aspirin and those who did not were equal, then the null hypothesis would be **accepted**. However, if those who took aspirin did have a reduced rate of heart attack within statistical significance, the null hypothesis would be **rejected**.

Degrees of Freedom - a value which must be determined in order to accept or reject the null hypothesis

Ex.

When rolling a die, the value of the **degrees of freedom** is 5 since there are 6 sides on a die.

Chi-square Equation:

Chi-square

Σ = summation

observed value

expected value

*It is important to remember that is the symbol for Chi-square. The value you calculate to be equal to does not need to be squared or have its square root calculated. (common mistake)

Critical Value Chart – when the Chi-square (x^2) value is compared to the values in the critical value chart, this comparison can be used to **accept** or **reject** the null hypothesis.

Ex.

Suppose in an experiment, there are 5 possible outcomes.

- Since there are 5 possible outcomes, the degrees of freedom are 4.
- With **4** degrees of freedom, the critical value will be 9.49
- If $x^2 >$ **9.49**, the null hypothesis will be **rejected**.
- If $x^2 <$ **9.49**, the null hypothesis will be **accepted**.

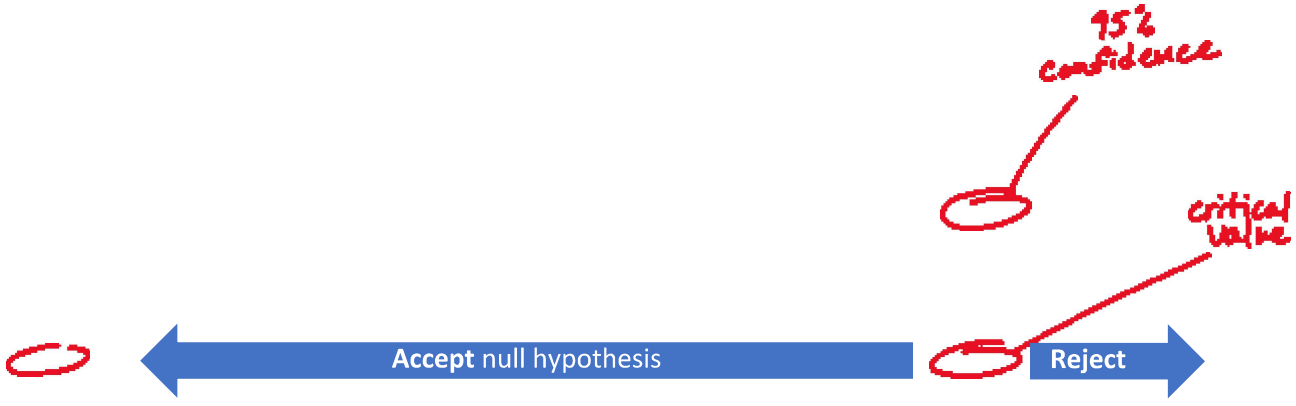

Table 5.3. Chi-square value

Degrees of Freedom	Probability								Non-significant	Signi-ficant	Highly significant
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
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
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21

Sample: A die is rolled 200 times and the observed and expected values are recorded below.

	⚀	⚁	⚂	⚃	⚄	⚅	**Total**
Observed (o)	**28**	**36**	**39**	**30**	**37**	**30**	**200**
Expected (e)	**33.33**	**33.33**	**33.33**	**33.33**	**33.33**	**33.33**	**200**

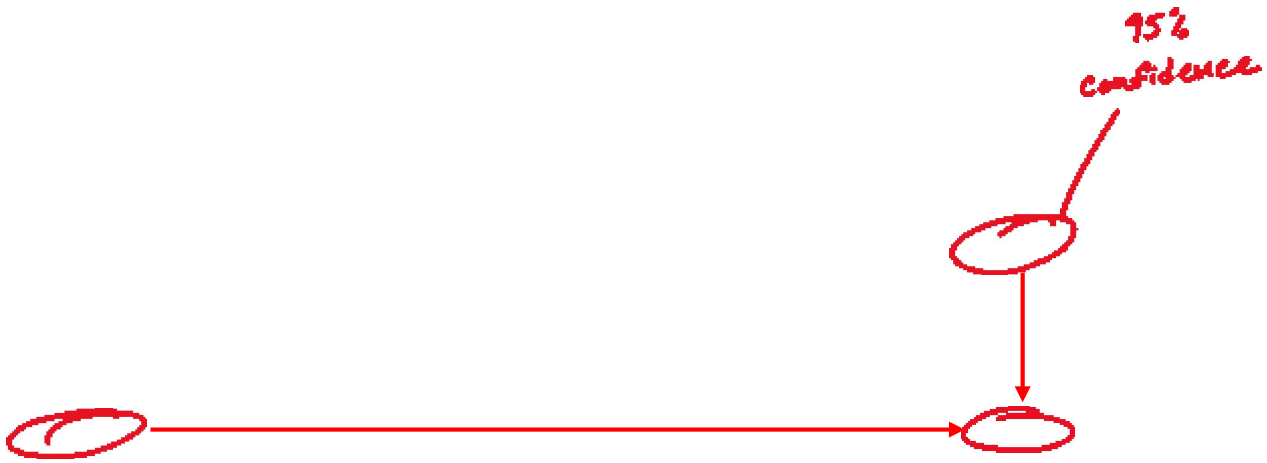

Table 5.3. Chi-square value

Degrees of Freedom	Probability						Non-significant		Signi-ficant	Highly significant
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
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
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21

Conclusion:

Since 3.08 is less than the critical value (11.07), the null hypothesis is **accepted**. Again, this means that although there is a slight difference between the observed and expected values, the disparity is not great enough to indicate that an external variable is responsible for the disparity. Instead, the difference between observed and expected values was simply due to chance.

Coin Flipping Activity:

Directions:

1. Gather a coin.
2. Create a **null** hypothesis, determine the **degrees of freedom**, and determine the **expected** probability of each outcome
3. Flip the coin 50 times and record the number of heads and tails **observed** values.
4. Calculate the Chi-square (x^2) value.
5. Compare the x^2 value to the critical value.
6. Either **accept** or **reject** your null hypothesis.
7. If your null hypothesis is rejected, consider possible causes for this disparity between the observed and expected outcomes.

Table 5.3. Chi-square value

Degrees of Freedom	Probability						Non-significant		Signi-ficant	Highly significant
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
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
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21