Who Uses Chi-Square Analysis?

Chi-square is most commonly used by researchers who are studying survey response data because it applies to categorical variables. Demography, consumer and marketing research, political science, and economics are all examples of this type of research.

Example

Let's say you want to know if gender has anything to do with political party preference. You poll 440 voters in a simple random sample to find out which political party they prefer. The results of the survey are shown in the table below:

	Republican	Democrat	Independent	Total
Male	100	70	30	200
Female	140	60	20	220
Total	240	130	50	440

To see if gender is linked to political party preference, perform a Chi-Square test of independence using the steps below.

Step 1: Define the Hypothesis

H0: There is no link between gender and political party preference.

H1: There is a link between gender and political party preference.

Step 2: Calculate the **Expected Values**

Now you will calculate the expected frequency.

$$\text{Expected Value} = \frac{(Row\ Total) * (Column\ Total)}{Total\ Number\ Of\ Observations}$$

For example, the expected value for Male Republicans is:

$$= \frac{(240) * (200)}{440} = 109$$

Similarly, you can calculate the expected value for each of the cells.

Expected Values				
	Republican	**Democrat**	**Independent**	**Total**
Male	109	59	22.72	200
Female	120	65	25	220
Total	240	130	50	440

Step 3: Calculate (O-E)2 / E for Each Cell in the Table

Now you will calculate the (O - E)2 / E for each cell in the table.

Where

O = Observed Value

E = Expected Value

$(O - E)^2/E$				
	Republican	**Democrat**	**Independent**	**Total**
Male	0.74311927	2.050847	2.332676056	200
Female	3.33333333	0.384615	1	220
Total	240	130	50	440

Step 4: Calculate the Test Statistic X2

X2 is the sum of all the values in the last table

= 0.743 + 2.05 + 2.33 + 3.33 + 0.384 + 1

= 9.837

Before you can conclude, you must first determine the critical statistic, which requires determining our degrees of freedom. The degrees of freedom in this case are equal to the table's number of columns minus one multiplied by the table's number of rows minus one, or (r-1) (c-1). We have (3-1)(2-1) = 2.

Finally, you compare our obtained statistic to the critical statistic found in the chi-square table. As you can see, for an alpha level of 0.05 and two degrees of freedom, the critical statistic is 5.991, which is less than our obtained statistic of 9.83. You can reject our null hypothesis because the critical statistic is higher than your obtained statistic.

This means you have sufficient evidence to say that there is an association between gender and political party preference.

Critical values of the Chi-square distribution with d degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	d	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

INTRODUCTION TO POPULATION GENETICS, Table D.1

x_1	x_2	y
1	1	1
1	0	0
0	1	0
0	0	0