

Statistical Test for the Difference Between Millennials' Preference for a Central City Location in 2015 and 2018

The Chi Squared is often used to compare two samples of discrete variables and conclude whether they have a statistically significant difference in distribution of outcomes. The test statistic, X^2 , is the squared sum of observed frequency of each response category minus the expected frequency all over the expected frequency. If this test statistic is greater than a critical value, then we can reject the idea hypothesis that these two samples have the same distribution of outcomes, and, rather, accept that they have a statistically significant difference in distribution of outcomes.

$$x^2 = \sum \frac{(O - E)^2}{E}$$

The critical value is found on a Chi Square table found using a desired significance level, α , and the degrees of freedom, which is the product of the number of groups minus one times the number of outcomes minus one. The formula for the test statistic falls below. The significance level used is 0.05 and the degrees of freedom are $(2-1)*(2-1) = 1$, so the critical value is $x^* = 3.841$.

Now, a table of observed outcomes will be constructed. There were $N_{2015} = 1,712$ respondents in the 2015 survey, and $p_{2015} = 12\%$ of them indicated they preferred living in the central city. In 2018, these numbers were $N_{2018} = 2,014$ respondents and $p_{2018} = 23\%$.

Observed	Prefer City	Don't	Total
2015	$N_{2015} * p_{2015}$	$N_{2015} * (1 - p_{2015})$	N_{2015}
2018	$N_{2018} * p_{2018}$	$N_{2018} * (1 - p_{2018})$	N_{2018}
Total	$N_p = \text{Sum of "Prefer City"}$	$N_d = \text{Sum of "Don't"}$	$N = N_{2015} + N_{2018}$

Now, this table will be filled out with the actual observations rounded to the closest integer.

Observed	Prefer City	Don't	Total
2015	$1712 * .12 = 205$	1507	1712
2018	463	1551	2014
Total	668	3058	3726

Now a similar table of the number of expected outcomes will be constructed. This is found for each cell by multiplying the column and row total together and dividing by N.

Expected	Prefer City	Don't	Total
2015	$N_{2015} * N_p$	$N_{2015} * N_d$	N_{2015}
2018	$N_{2018} * N_p$	$N_{2018} * N_d$	N_{2018}
Total	$N_p = \text{Sum of "Prefer City"}$	$N_d = \text{Sum of "Don't"}$	$N = N_{2015} + N_{2018}$

Now, the table will be filled with the actual expected values.

Expected	Prefer City	Don't	Total
2015	$668 * 1712 / 3726 = 306.93$	1405.7	1712
2018	361.07	1652.93	2014
Total	668	3058	3726

Now, the test statistic can be calculated using as $\chi^2 = \sum \frac{(O - E)^2}{E}$ detailed below:

$$\chi^2 = \frac{(205 - 306.93)^2}{3726} + \frac{(1507 - 1405.7)^2}{3726} \dots = 76.303$$

Because our test statistic is greater than the critical value of 3.841, we can conclude that these two samples, the group of Millennial homebuyers sampled in 2015 and in 2018, have a statistically significant difference in distribution of outcomes when it comes to their preference of living in the central city. What this means is that the change in preferences between 2015 and 2018 for Millennials is truly reflective of a significant change in what they desire instead of just sampling "noise".