Test of Homogeneity

A ***test of homogeneity*** involves testing the null hypothesis that the proportions of elements with certain characteristics in two or more different populations are the same against the alternative hypothesis that these proportions are not the same.

Example: Suppose you are interested in knowing whether the distribution of income classes (low, middle, high) are the same for males and for females in your city. You take a random sample of 500 males and 450 females from your state; you determine the income level of each individual. The data are summarized below in the two-way table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low Income | Middle Income | High Income |
| Male | 109 | 365 | 26 |
| Female | 192 | 249 | 9 |

Is the distribution of income the same or different for males and females?

Step 1: The distribution of income is not the same for males and females

Step 2: $H\_{0}: the distribution \left(proportion\right) of income under each distribution is the same for males and females$

$H\_{1}: the distribution \left(proportion\right) of income under each distribution is not the same for males and females$

Step 3: Test of homogeneity is always chi-squared distributed. This is also always a right-tailed test.

Step 4: Calculate the expected frequencies to derive the test statistic

To calculate the expected frequencies, compute the row and column totals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Low Income | Middle Income | High Income | Total |
| Male | 109 | 365 | 26 | 500 |
| Female | 192 | 249 | 9 | 450 |
| Total | 301 | 614 | 35 | **950** |

Expected Frequencies are calculates the same way as before:



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Low Income | Middle Income | High Income | Total |
| Male | $$\frac{500×301}{950}$$ | $$\frac{500×614}{950}$$ | $$\frac{500×35}{950}$$ | 500 |
| Female | $$\frac{450×301}{950}$$ | $$\frac{450×614}{950}$$ | $$\frac{450×35}{950}$$ | 450 |
| Total | 301 | 614 | 35 | **950** |

To calculate the test statistic use the following formula:



Conduct a right-tailed rejection region approach. The degrees of freedom is $df=(R-1)(C-1)$.