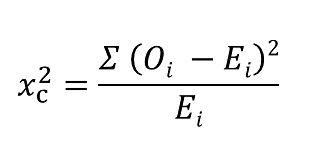
.

Aim: Performing the Hypothesis testing using the Chi-squared test

The table below is a survey response to 4 categorical variables: people in categories from 18–29, 30–44, 45–64 and >65 years, and their movie genre inclination, which is “Action/Adventure”, “Romance” and “Biography”. Is there any evidence of a relationship between the age group and their movie genre inclination, at 5% significance level?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Action/Adventure | Romance | Biography | Total |
| 18-29 | 141 | 68 | 4 | 213 |
| 30-44 | 179 | 159 | 7 | 345 |
| 45-64 | 220 | 216 | 4 | 440 |
| 65&older | 86 | 101 | 4 | 191 |
| Total | 626 | 544 | 19 | 1189 |



Where

c = Degrees of freedom

O = Observed Value

E = Expected Value

Step 1: Define the Hypothesis

H0: There is no link between age group and movie genre

H1: There is a link between age group and movie genre

Step 2: Calculate the Expected Values

calculating the expected frequency.

https://www.simplilearn.com/ice9/free_resources_article_thumb/Chi_Sq_formula_1.png

**EXPECTED VALUES**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Action/Adventure | Romance | Biography | Total |
| 18-29 | **112.4** | **97.45** | **3.43** | 213 |
| 30-44 | **181.64** | **157.84** | **5.51** | 345 |
| 45-64 | **231.65** | **201.31** | **7.03** | 440 |
| 65&older | **100.56** | **87.38** | **3.05** | 191 |
| Total | 626 | 544 | 19 | 1189 |

### 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.

O = Observed Value

E = Expected Value

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Action/Adventure | Romance | Biography | Total |
| 18-29 | 7.42 | 8.90 | 0.094 | 213 |
| 30-44 | 0.038 | 0.0085 | 0.403 | 345 |
| 45-64 | 0.586 | 1.07 | 1.307 | 440 |
| 65&older | 2.10 | 2.122 | 0.296 | 191 |
| Total | 626 | 544 | 19 | 1189 |

### Step 4: Calculate the Test Statistic X2

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

**=7.42+8.90+0.094+0.038+0.0085+0.403+0.586+1.07+1.307+0.296+2.10+2.122**

**=24.345**

Before we conclude, we 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 (4-1)(3-1) = 6

Finally, we compare our obtained statistic to the critical statistic found in the chi-square table. As we can see, for an alpha level of 0.05 and 6 degrees of freedom, the critical statistic is 12.592, which is less than our obtained statistic of 24.345. we can reject our null hypothesis because the critical statistic is higher than your obtained statistic.

This means we have sufficient evidence to say that there is an association between age group and movie genre.

