

Worksheet: Chi-square test of independence

The Scene: We ask a sample of Linfield students the following question:

If you could choose one of the following accomplishments for your life, which would you choose:

- To win an Olympic gold medal
- To win a Nobel Prize
- To win an Academy Award
- To become President of the United States

Is a student's answer to this question independent of where at Linfield they intend to earn their major (School of Nursing, School of Business, or The College of Arts and Sciences)?

We conduct a chi-square test of independence on these hypotheses:

H_0 : There is no association between the location of a person's intended major at Linfield and their answer to the achievement question.

H_a : There is an association between these categorical variables.

- Record in the two-way table below the results of the class survey (conducted via google form), and fill in the row and column sums as well.

observed counts	gold	nobel	oscar	pres	sum
business					
cas					
nursing					
sum					

- Overall, what percentage of students in the survey chose gold medal as their preferred achievement?
- Overall, how many students indicated business as their likely major?
- If the same percentage of business students chose gold medal as was the case for all students, how many business students would have chosen gold medal?
- Under the assumption that the null hypothesis is true, namely that there is no association between these two categorical variables, then the expected count for cell (row i , column j) is given by the formula

$$E_{i,j} = \frac{(\text{row } i \text{ total}) \cdot (\text{column } j \text{ total})}{\text{overall total}}.$$

Using this formula determine the expected counts for each cell.

expected counts	gold	nobel	oscar	pres	sum
business					
cas					
nursing					
sum					

6. Observe that the expected count in the cell (business, gold) should match your answer to problem [4]. Does it?

7. Determine the chi-square score,

$$\chi^2 = \sum_{\text{all cells}} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}.$$

8. In theory, this chi-square score lives in a chi-square distribution with how many degrees of freedom?

9. Use the `pchisq()` command in R to determine the p-value for this test.

10. If there is no association between a student's intended major and their chosen achievement in the survey question, then how likely would it be to gather data that produced a chi-square score as large or larger than the one we computed in [9]?

11. Based on your analysis, do you reject H_o in favor of H_a , or do you fail to reject H_o ? Explain in a sentence.