Multi category chi-squared tests: Takeaways

Syntax

- The `chisq.test()` function takes a matrix of values and returns three things: the chi-squared test statistic, the degrees of freedom and the resulting p-value

```
library(readr)
income <- read.csv("income.csv")
data <- table(income$sex,    income$high_income)
chisq.test(data)
>>

    Pearson's   Chi-squared  test  with  Yates'   continuity   correction

data:   data
X-squared = 1517.8,   df = 1,  p-value < 2.2e-16
```

Concepts

- In a multiple category chi-squared test, we calculate expected values across our whole dataset.
- We can calculate the chi-squared value by using the following steps:
  - Subtract the expected value from the observed value.
  - Subtract the difference.
  - Divide the squared difference by the expected value.
  - Repeat for all observed and expected values and add up all the values.

- Formula for chi-squared:

- Finding that a result isn't significant doesn't mean that no association between the columns exists. Finding a statistically significant result doesn't imply anything about what the correlation is.
- Chi-squared tests can only be applied in the case where each possibility within a category is independent.

Resources

- [Chi-squared test of association](#)