chisq_independence_test

October 29, 2020

1 Using Pearson's χ^2 Test of Independence

2 1. Pearson's χ^2 (Independence) Test

We can also use Pearson's χ^2 to solve a different sort of problem. In particular, we can use Pearson's χ^2 to test the extent to which two categorical variables are independent.

3 1.1 Pearson's χ^2 (Independence) Test Example

Suppose we would like to teach cats to dance.

We have two training systems: using food as a reward, and using affection as a reward. Suppose after a week of training the cats, we test dancing ability. So, we have two categorical variables: *training* and *dance*, each with two levels.

.	.	Food as reward	Affection as reward
Cat Dances?	Yes	28	48
.	No	10	114

From these data, are the *training* and *dance* varialbes independent?

*Source: Field *et al.* (2012)

3.0.1 1.1.1 Pearson's χ^2 Independence Test (cont.)

The test statistic is χ^2 and is computed using:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}},$$

where

$$E_{i,j} = \frac{\text{row-total}_i \times \text{column-total}_j}{N}$$

and where $O_{i,j}$ is the observed count in cell i, j and $E_{i,j}$ is the expected count for cell i, j under the null hypothesis.

3.0.2 1.1.2 Pearson's χ^2 Independence Test (cont.)

Note: - Degrees of freedom: $ df = (r - 1)(c - 1) $ where r is the number of rows, and c is the number of columns - Assumption that observations are independent from one another + E.g., In above example, a cat could only be in one *training* condition

4 2. Pearson's χ^2 Independence Test in R

```
In [3]: can_dance <- c(rep(TRUE, 76), rep(FALSE, 124))

        training <- c(rep("food", 28), rep("affection", 48), rep("food", 10), rep("affection",

        cats <- data.frame(can_dance, training)

        head(cats)
```

A data.frame: 6 Œ 2

	can_dance <lgl>	training <chr>
1	TRUE	food
2	TRUE	food
3	TRUE	food
4	TRUE	food
5	TRUE	food
6	TRUE	food

4.1 2.1 Running χ^2 Test of Independence

```
In [4]: # sanity check to make sure data are correct
        xtab1 <- table(cats$can_dance, cats$training)

        print(xtab1)


        affection food
  FALSE       114   10
  TRUE         48   28


In [5]: test1 <- chisq.test(cats$training, cats$can_dance)

        print(test1)


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

data:  cats$training and cats$can_dance
X-squared = 23.52, df = 1, p-value = 1.236e-06
```