

Test 03: Categorical Y vs categorical X

Richard

1/30/2022

Load required packages

```
library(tidyverse)
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
```

```
## v tibble  3.1.1      v dplyr  1.0.5
```

```
## v tidyr   1.1.3      v stringr 1.4.0
```

```
## v readr   1.4.0      v forcats 0.5.1
```

```
# if you're using macOS, you can run: library(dplyr)
```

```
library(skimr)
```

Prepare Data

```
Hsb <- within(  
  read.csv("https://stats.idre.ucla.edu/stat/data/hsb2.csv")  
  race <- as.factor(race)  
  schtyp <- as.factor(schtyp)  
  prog <- as.factor(prog)  
)
```

Categorical Y and categorical X

For example, we use the hsb2 data file, let's see if there is a relationship between the type of school attended (schtyp) and students' gender (female).

Frequency table

- ▶ The below code show the frequency table between gender and school type:

```
Hsb %>% count(female, schtyp)
```

```
##   female schtyp  n
## 1      0      1 77
## 2      0      2 14
## 3      1      1 91
## 4      1      2 18
```

- ▶ Can you explain this data?

A matrix format

- ▶ Each row is gender
- ▶ Each column is each school type

```
table(Hsb$female, Hsb$schtyp)
```

```
##  
##      1  2  
##  0 77 14  
##  1 91 18
```

Chi-square test

```
chisq.test(table(Hsb$female, Hsb$schtyp))
```

```
##
```

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

```
##
```

```
## data:  table(Hsb$female, Hsb$schtyp)
```

```
## X-squared = 0.00054009, df = 1, p-value = 0.9815
```

Discussion

- ▶ These results indicate that there is no statistically significant relationship between the type of school attended and gender (chi-square with one degree of freedom = 0.0005, $p = 0.98$).

Another example

Please check the relationship between gender (female) and socio-economic status (ses) using chi-square test.

Assumption of chi-square test

- ▶ Remember that the chi-square test assumes the expected value of each cell is five or higher.
- ▶ However, if this assumption is not met in your data, please see the section on Fisher's exact test below.

Fisher's exact test

```
fisher.test(table(Hsb$female, Hsb$schtyp))
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data: table(Hsb$female, Hsb$schtyp)
```

```
## p-value = 0.8492
```

```
## alternative hypothesis: true odds ratio is not equal to
```

```
## 95 percent confidence interval:
```

```
## 0.4755259 2.5307479
```

```
## sample estimates:
```

```
## odds ratio
```

```
## 1.087428
```

Practice

Please check the relationship between race (race) and school type (schtyp)

Quiz time

Hmm...