Contents

Chi-Square Test of Independence

The chi-square test of independence is used to analyze the frequency table (i.e. contingency table) formed by two categorical variables. The chi-square test evaluates whether there is a significant association between the categories of the two variables.

```
 #Import the data
file_path <- "http://www.sthda.com/sthda/RDoc/data/housetasks.txt"
housetasks <- read.delim(file_path, row.names = 1)
head(housetasks)
```

```
##            Wife Alternating Husband Jointly
## Laundry     156           14       2       4
## Main_meal   124           20       5       4
## Dinner       77           11       7      13
## Breakfeast   82           36      15       7
## Tidying      53           11       1      57
## Dishes       32           24       4      53
```

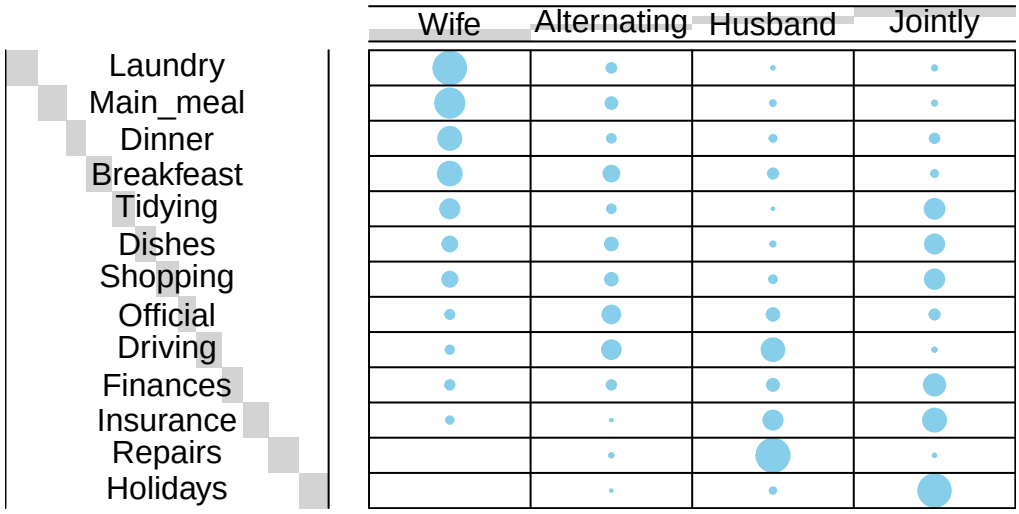
```
The **chi-square test of independence** is used to analyze the frequency table (i.e. contengency table)
```

The data is a contingency table containing 13 housetasks and their distribution in the couple: * **rows are the different tasks** * **values are the frequencies of the tasks done : ** * **by the wife only** * *alternatively* * **by the husband only** * **or jointly**

Graphical display of contengency tables

Contingency table can be visualized using the function **balloonplot()** [in `gplots` package]. This function draws a graphical matrix where each cell contains a dot whose size reflects the relative magnitude of the corresponding component.

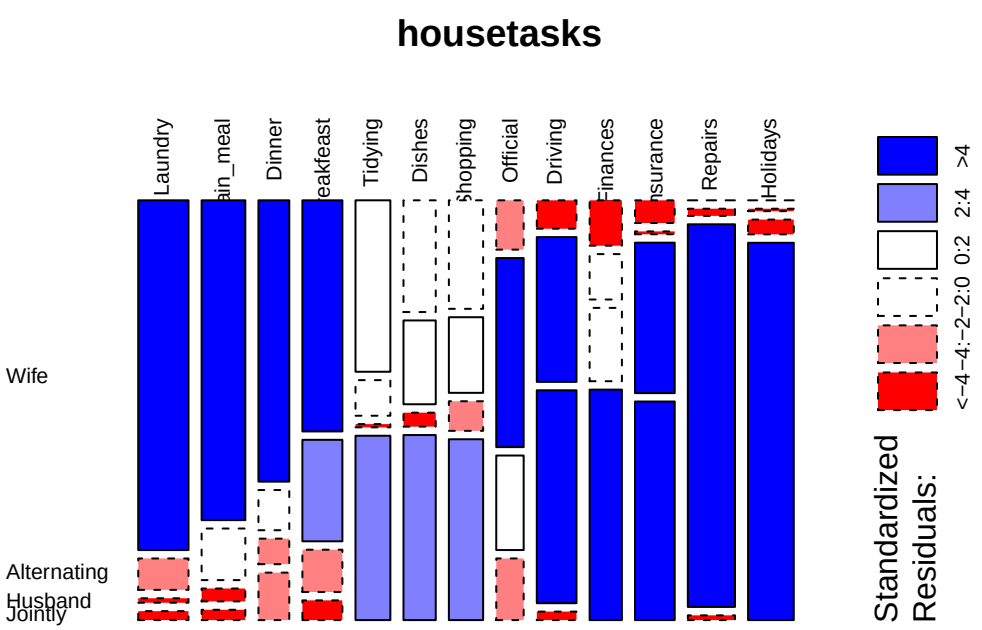
```
library("gplots")
# 1. convert the data as a table
dt <- as.table(as.matrix(housetasks))
# 2. Graph
balloonplot(t(dt), main ="housetasks", xlab ="", ylab="",
            label = FALSE, show.margins = FALSE)
```




```
Note that, row and column sums are printed by default in the bottom and right margins, respectively. Th
```

It's also possible to visualize a contingency table as a mosaic plot. This is done using the function mosaicplot() from the built-in R package garphics:

```
library("graphics")
mosaicplot(dt, shade = TRUE, las=2,
           main = "housetasks")
```



- The argument shade is used to color the graph
- The argument las = 2 produces vertical labels
- **Note that the surface of an element of the mosaic reflects the relative magnitude of its value.**
- Blue color indicates that the observed value is higher than the expected value if the data were random
- Red color specifies that the observed value is lower than the expected value if the data were random
- From this mosaic plot, it can be seen that the house tasks Laundry, Main_meal, Dinner and breakfast (blue color) are mainly done by the wife in our example.

There is another package named vcd, which can be used to make a mosaic plot (function mosaic()) or an association plot (function `assoc()`).

Chi-square test basics

Chi-square test examines whether rows and columns of a contingency table are statistically significantly associated.

- **Null hypothesis (H0):** the row and the column variables of the contingency table are independent.
- **Alternative hypothesis (H1):** row and column variables are dependent For each cell of the table, we have to calculate the expected value under null hypothesis.

```
chisq <- chisq.test(housetasks)
chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  housetasks
## X-squared = 1944.5, df = 36, p-value < 2.2e-16
```

- In our example, the row and the column variables are statistically significantly associated (p-value = 0).

The observed and the expected counts can be extracted from the result of the test as follows:

```
# Observed counts
chisq$observed
```

```
##            Wife Alternating Husband Jointly
## Laundry     156          14       2       4
## Main_meal   124          20       5       4
## Dinner       77          11       7      13
## Breakfeast   82          36      15       7
## Tidying      53          11       1      57
## Dishes       32          24       4      53
## Shopping     33          23       9      55
## Official     12          46      23      15
## Driving      10          51      75       3
## Finances     13          13      21      66
## Insurance     8           1      53      77
## Repairs       0           3     160       2
## Holidays      0           1       6     153
```

```
# Expected counts
round(chisq$expected,2)
```

```
##             Wife Alternating Husband Jointly
## Laundry    60.55       25.63   38.45   51.37
## Main_meal  52.64       22.28   33.42   44.65
## Dinner     37.16       15.73   23.59   31.52
## Breakfeast 48.17       20.39   30.58   40.86
## Tidying    41.97       17.77   26.65   35.61
## Dishes     38.88       16.46   24.69   32.98
## Shopping   41.28       17.48   26.22   35.02
## Official   33.03       13.98   20.97   28.02
## Driving    47.82       20.24   30.37   40.57
## Finances   38.88       16.46   24.69   32.98
## Insurance  47.82       20.24   30.37   40.57
## Repairs    56.77       24.03   36.05   48.16
## Holidays   55.05       23.30   34.95   46.70
```