STAT 477/577 - Technology Guide

Module 2 - Section 5

Measures of Association

Below is an explanation of the R commands and functions needed to calculate measures of association in a contingency table.

- **ϕ Coefficient**

  Base R does not contain a function to calculate the ϕ coefficient (r_ϕ). I wrote a function called `r.phi` to calculate this value. To obtain the value from the lecture notes using the modified data on parent and student smoking, we will need to first read in the data file `smokingc.csv`.

```
smokingc.data<- read.csv(file.choose(), header = T)
```

  Now, we will order the categories for each variable.

```
smokingc.data$Student<- factor(smokingc.data$Student,
                          levels = c("Non-Smoker", "Smoker"))
smokingc.data$Parent<- factor(smokingc.data$Parent,
                          levels = c("Neither", "One or Both"))
```

  The contingency table can be calculated as:

```
smokingc.table<- table(smokingc.data$Parent,
                       smokingc.data$Student)
```

  and the ϕ coefficient value is

```
r.phi(smokingc.table)

##  X-squared
## 0.07175744
```

- **Cramer's V**

  Base R does not contain a function to calculate Cramer's V ($\hat{\phi}_C$). I wrote a function called `phi.c` to calculate this value. To obtain the value from the lecture notes using the data on parent and student smoking, we will need to first read in the data file `smoking.csv`.

```
smoking.data<- read.csv(file.choose(), header = T)
```

Now, we will order the categories for each variable.

```
smoking.data$Student<- factor(smoking.data$Student,
                          levels = c("Non-Smoker", "Smoker"))
smoking.data$Parent<- factor(smoking.data$Parent,
                          levels = c("Neither", "One", "Both"))
```

The contingency table can be calculated as:

```
smoking.table<- table(smoking.data$Parent,
                      smoking.data$Student)
```

and Cramer V is:

```
phi.c(smoking.table)

##  X-squared
## 0.08360076
```

- **Goodman-Kruskal Gamma**

  The Goodman-Kruskal Gamma statistic is used for assessing the level of directional association between two ordinal variables. A function (`GKgamma`) to calculate this statistic for an I by J contingency table is given in the R package `vcdExtra`. We still begin by reading in the data file `jobdata.csv`.

```
jobs.data<- read.csv(file.choose(), header = T)
```

We will then order the categories for both variables.

```
jobs.data$Physically<- factor(jobs.data$Physically,
                   levels = c("Seldom", "Sometimes", "Usually"))
jobs.data$Psychologically<- factor(jobs.data$Psychologically,
                   levels = c("Seldom", "Sometimes", "Usually"))
```

Now, we can calculate the contingency table

```
jobs.table<- table(jobs.data$Physically,
                   jobs.data$Psychologically)
```

and then use the function `GKgamma` to calculate the Goodman-Kruskal Gamma statistic and obtain a confidence interval.

```
GKgamma(jobs.table)

## gamma         : 0.236
## std. error    : 0.038
## CI            : 0.16 0.311
```