

Social Robotics

B. Nebel, F. Lindner, L. Wächter
Summer Semester 2019

University of Freiburg
Department of Computer Science

Exercise Sheet 5

Due: June 17, 2019

Exercise 5.1 (Chi-squared Test of Independence)

We will use a built-in dataset in order to compare the relation of lifestyle aspects of students.

- Load the library `MASS` to use the built-in dataset `survey`.
- We are interested in the smoking (`Smoke`) and exercise (`Exer`) habits of the students. Create a table out of them for a better overview.
- Apply the `chisq.test()` function.
- What can you see and interpret?
- As you see, the function also produces a warning. This is based on the small cell values in the data. Dependent on your research goal and hypothesis you can try to rearrange your data (into larger groups) in order to get more stable results. In this example it would be possible to
 - combine all people who do not exercise frequently/ do exercise at all
 - combine all people who smoke more than a threshold you choose.

Decide for one option and give an argument for it.

- Change your data according to your decision
- Calculate the chi-test again.
- What can you see and interpret now?

Exercise 5.2 (Chi-Squared Goodness-Of-Fit Test)

We will create some fake data and check whether to check whether the national distribution of musicians is a good fit for the sample city's one.

```
national_music
classical jazz  hiphop  metal  pop
15000     50000  60000   35000  100000
sample_city_music
classical jazz  hiphop  metal  pop
75       250   300     150   600
```

- Create two vectors (`national_music` and `sample_city_music`) for the number of professional musicians in different genres using the `rep` function.
- Create a table of `sample_city_music`. This is equivalent to your observed data.
- Use the function `prop.table` to calculate the national ratios.
- Use the function `chisq.test()` to do the chi-squared test of independence.
- What conclusions can you draw?