Assignment 3: Chi-square test

Sanjiv Kumar

11/28/2021

Part 1. Personality and animals

Introduction

A group of students were surveyed in terms of personality, depending upon preferred animal chosen, to identify if there are any significant association between the personality of the student and preferred animal. The data used here was provided with the assignment (Ref 1).

Method

The data was imported using `read.table` and converted to tibble using `as_tibble`. The contingency table was prepared using `addmargins(table(personality_animal))`. Chi-square test was performed to identify any significant association between the personality type using `chisq.test(personality_animal_table, correct = TRUE)`.

Result

Examining Data Following is the distribution of the animal preference as per the personality type (Table 1).

Table 1: Personality and animals
Animal preference as per the personality type

Personality	Animal	n
Extrovert	Bat	16
Introvert	Bat	11
Extrovert	Rabbit	13
Introvert	Rabbit	18
Extrovert	Tiger	20
Introvert	Tiger	2
Extrovert	Whale	10
Introvert	Whale	10

Contingency Table From the data provided (Ref 1) a contingency table was prepared with the sums at the margins (Table 2).

Table 2: Contingency Table

	Bat	Rabbit	Tiger	Whale	Sum
Extrovert	16	13	20	10	59
Introvert	11	18	2	10	41
Sum	27	31	22	20	100

χ^2 Chi-square test

```
##
##  Pearson's Chi-squared test
##
## data:  personality_animal_table
## X-squared = 13.662, df = 3, p-value = 0.003403
```

- Type of test: The χ^2 test for associations
- Chi-square value: 13.6623091
- df: 3
- p-value: 0.0034028

Fisher's exact test Since the smallest count is less then 5, i.e. 2 (see Table 2), we want to use Fisher's exact test.

```
##
##  Fisher's Exact Test for Count Data
##
## data:  personality_animal_table
## p-value = 0.001897
## alternative hypothesis: two.sided
```

- Type of test: Fisher's Exact Test for Count Data
- p-value: 0.0018966

Data visualization

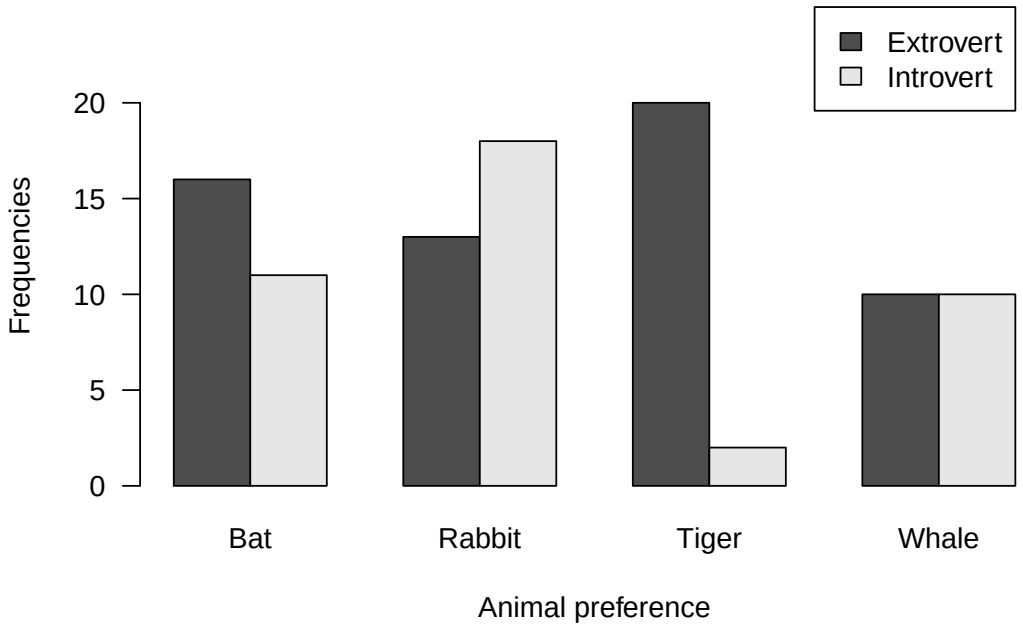

Figure 1: Animal preference as per the personality type.

Discussion

Here association between personality type and animal preferences were examined.

Here hypotheses are:

- H0: There is no association between personality type and animal preference.
- H1: There is significant association between personality type and animal preference.

From the χ^2 test for associations, the p-value: 0.0034028 having df: 3 hence $p<0.05$, therefore, we reject the H0 (null hypothesis) and we conclude that there is significant association between personality type and animal preference. The results of `Fisher's exact test` shows the p-value: 0.0018966 and therefore $p<0.05$, leading to similar conclusion i.e. rejecting H0 and thus, shows significant association between personality type and animal preference.

References

1. Data provided with the assignment as file `Personality animal.txt`.

Part 2. Chi-square analysis of a chosen subject

Introduction

The current data is a subset of the dataset (Ref 1) collected for a food app with respect to the eating habits of the people in Stockholm, Sweden in 2020 (may be heavily influenced by ongoing COVID-19). Here only the columns Gender (Male and Female) and frequency of Eating Outside (Daily, Monthly, Weekly and Yearly) is considered for the analysis. With this data we are trying to answer if there's an effect of gender on frequency of eating outside.

- H0: There is no association between gender and frequency of eating outside.
- H1: There is significant association between gender and frequency of eating outside.

Method

The data was imported using `read.csv`. The contingency table was prepared using `addmargins(table(eating_outside))`. Chi-square test was performed to identify any significant association between the personality type using `Fisher's exact test` as there were values less then 5, however `chisq.test(eating_outside, correct = TRUE)` was also performed for curiosity.

Result

Examining Data Following is the frequency distribution of eating outside based upon gender (Table 1).

Table 1: Frequency of eating outside
Influence of gender

Gender	EatingOutside	n
Female	Daily	0
Male	Daily	4
Female	Monthly	23
Male	Monthly	14
Female	Weekly	11
Male	Weekly	11
Female	Yearly	3
Male	Yearly	2

Contingency Table From the data provided (Ref 1) a contingency table was prepared with the sums at the margins (Table 2).

Table 2: Contingency Table

	Daily	Monthly	Weekly	Yearly	Sum
Female	0	23	11	3	37
Male	4	14	11	2	31
Sum	4	37	22	5	68

χ^2 Chi-square test

```
##
##  Pearson's Chi-squared test
##
## data:  eating_outside_table
## X-squared = 5.9058, df = 3, p-value = 0.1163
```

- Type of test: The χ^2 test for associations

- Chi-square value: 5.9057565
- df: 3
- p-value: 0.1162865

Fisher's exact test Since the smallest count is less then 5, i.e. (0, 4, 3, 2, see Table 2), we need to use Fisher's exact test. In this case χ^2 Chi-square test may yield incorrect values.

```
##
##  Fisher's Exact Test for Count Data
##
## data:  eating_outside_table
## p-value = 0.1178
## alternative hypothesis: two.sided
```

- Type of test: Fisher's Exact Test for Count Data
- p-value: 0.1178369

Data visualization

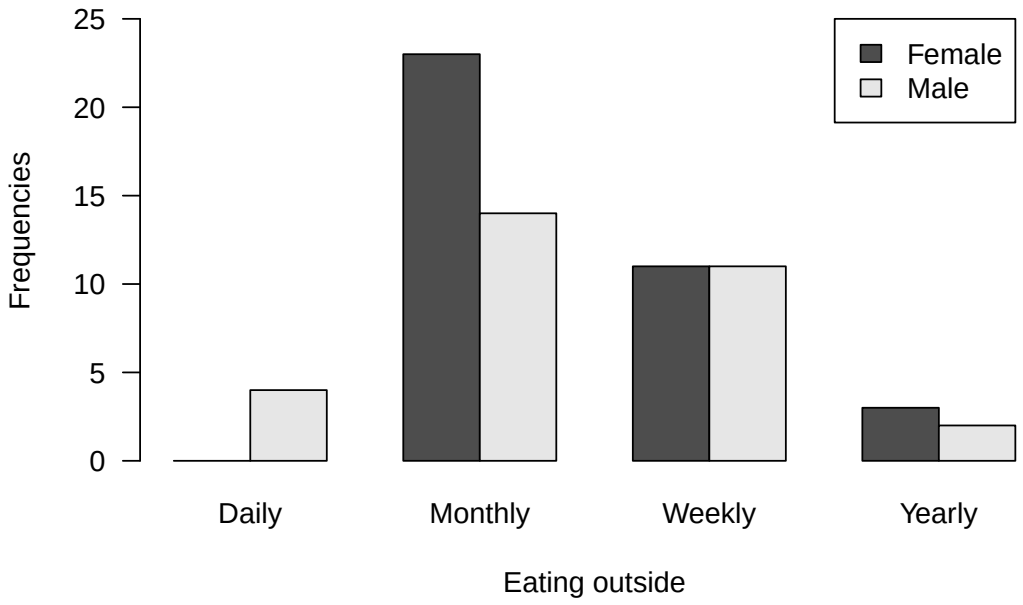

Figure 1: Influence of gender on frequency of eating outside

Discussion

Using a subset of the data collected to study the if there is association of gender with the frequency of eating outside, we formulated following hypothesis.

- H0: There is no association between gender and frequency of eating outside.
- H1: There is significant association between gender and frequency of eating outside.

Since there were values (frequencies) less then 5 in the contingency table, we conducted Fisher's exact test which shows that the p-value: 0.1178369 i.e. p>0.05 and therefore, we accept the H0 i.e. there is no association between gender and frequency of eating outside from this data.

Just for the sake of curiosity we also performed χ^2 test for associations, which resulted in a warning "Chi-squared approximation may be incorrect", and the p-value: 0.1162865, which is also p>0.05 and thus leading to similar conclusion i.e. accepting H0 and thus, show that there is no association between the gender and frequency of eating outisde based on this dataset.

References

1. Data is collected for a private food app based company, I can upload the data after removing crucial/specific details if required, please let me know.