DEVI AHILYA VISHWAVIDYALAYA

SCHOOL OF STATISTICS

NAME : VANSHIKA DHAKAD

ROLL NO : ST4A2108

SUBJECT : R-PROGRAMMING ASSIGNMENT

CLASS : B.Sc.(Hons)-ASA

SEMESTER : 4^{th}

SUBMITTED TO : CHANDRESH SIR

Case Study – Effectiveness of a drug treatment

To test the effectiveness of a drug for a certain medical condition, we will consider a hypothetical case.

Suppose we have 105 patients under study and 50 of them were treated with the drug. Moreover, the remaining 55 patients were kept under control samples. Thus, the health condition of all patients was checked after a week.

With the following table, we can assess if their condition has improved or not. By observing this table, one can you tell if the drug had a positive effect on the patient?

Here in this example, we can see that 35 out of the 50 patients showed improvement. Suppose if the drug had no effect, the 50 will split the same proportion of the patients who were not given the treatment. Here, in this case, improvement of the control case is high as about 70% of patients showed improvement, since both categorical variables which we have already defined must have only 2 levels. Also, it was sort of perceptive today that the drug treatment and health condition are dependent.

id	treatment	Improvement
1	treated	Improved
2	treated	Improved
3	not-treated	Improved
4	treated	Improved
5	treated	not-improved
6	treated	not-improved
7	not-treated	not-improved
8	treated	not-improved
9	not-treated	Improved
10	treated	Improved
11	not-treated	Improved
12	not-treated	not-improved
13	not-treated	not-improved
14	not-treated	not-improved
15	not-treated	Improved
16	not-treated	Improved
17	treated	Improved
18	treated	Improved
19	not-treated	not-improved
20	not-treated	not-improved
21	treated	not-improved
22	not-treated	not-improved
23	treated	not-improved
24	not-treated	Improved
25	treated	Improved
26	treated	Improved
27	not-treated	not-improved
28	not-treated	Improved
29	treated	not-improved
30	treated	Improved
31	not-treated	not-improved
32	not-treated	not-improved
33	treated	Improved
34	not-treated	Improved
35	treated	not-improved
36	not-treated	Improved
37	treated	Improved
38	not-treated	not-improved
39	not-treated	Improved
40	treated	Improved

41	not-treated	Improved
42	not-treated	Improved
43	not-treated	not-improved
44	not-treated	Improved
45	not-treated	Improved
46	treated	Improved
47	treated	not-improved
48	not-treated	not-improved
49	treated	Improved
50	treated	Improved
51	not-treated	not-improved
52	treated	Improved
53	not-treated	Improved
54	treated	Improved
55	treated	Improved
56	not-treated	Improved
57	treated	Improved
58	not-treated	not-improved
59	treated	Improved
60	treated	Improved
61	treated	Improved
62	not-treated	Improved
63	treated	not-improved
64	treated	not-improved
65	not-treated	Improved
66	not-treated	Improved
67	not-treated	Improved
68	not-treated	not-improved
69	not-treated	not-improved
70	treated	Improved
71	treated	not-improved
72	not-treated	not-improved
73	treated	not-improved
74	not-treated	Improved
75	not-treated	not-improved
76	not-treated	not-improved
77	treated	not-improved
78	not-treated	Improved
79	treated	Improved
80	treated	Improved
81	treated	Improved

82	not-treated	not-improved
83	treated	Improved
84	not-treated	not-improved
85	treated	Improved
86	not-treated	Improved
87	not-treated	not-improved
88	treated	Improved
89	not-treated	not-improved
90	treated	Improved
91	not-treated	not-improved
92	not-treated	Improved
93	treated	not-improved
94	treated	not-improved
95	not-treated	not-improved
96	treated	Improved
97	not-treated	Improved
98	treated	Improved
99	not-treated	not-improved
100	not-treated	Improved
101	treated	Improved
102	treated	Improved
103	not-treated	not-improved
104	treated	Improved
105	not-treated	not-improved

Chi-Square Test

Particularly in this test, we have to check the p-values. Moreover, like all statistical tests, we assume this test as a null hypothesis and an alternate hypothesis.

The main thing is, we reject the null hypothesis if the p-value that comes out in the result is less than a predetermined significance level, which is 0.05 usually, then we reject the null hypothesis.
H0: The two variables are independent.
H1: The two variables relate to each other.

In the case of a null hypothesis, a chi-square test is to test the two variables that are independent.

Syntax of a chi-square test:
chisq.test(data)

Following is the description of the chi-square test parameters:

- The input data is in the form of a table that contains the count value of the variables in the observation.
- We use chisq.test function to perform the chi-square test of independence in the native stats package in R. For this test, the function requires the contingency table to be in the form of a matrix. Depending on the form of the data, to begin with, this can need an extra step, either combining vectors into a matrix or cross-tabulating the counts among factors in a data frame.
- We use read.table and as.matrix to read a table as a matrix. While using this, be careful of extra spaces at the end of lines. Also, for extraneous characters on the table, as these can cause errors.

R Code

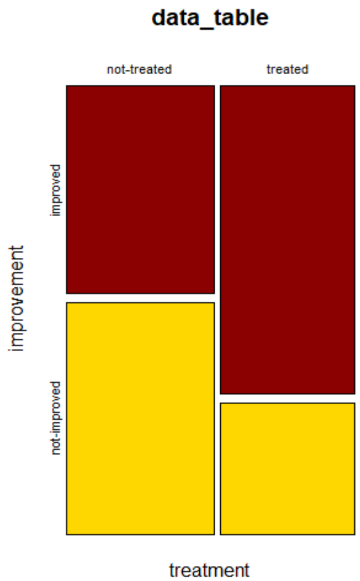
We will work on R by doing a chi-squared test on the treatment (X) and improvement (Y) columns in treatment.csv

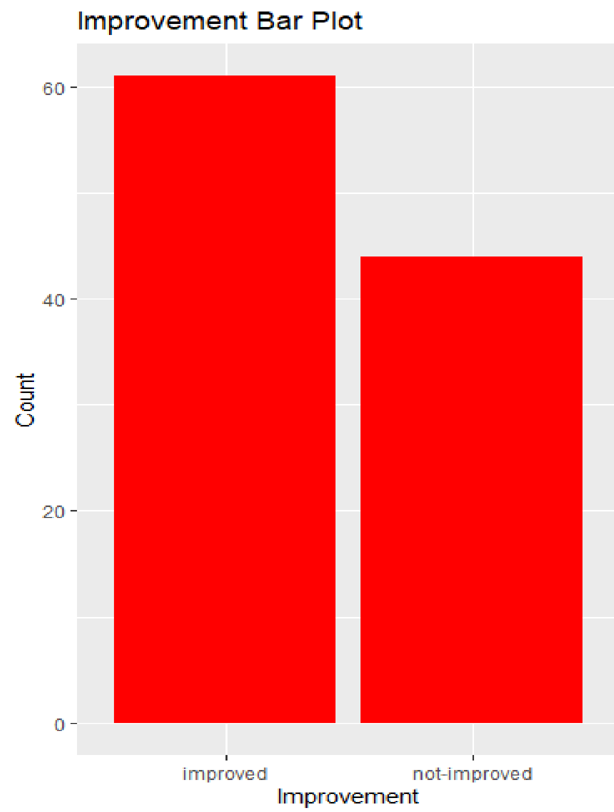
First, read in the treatment.csv data.

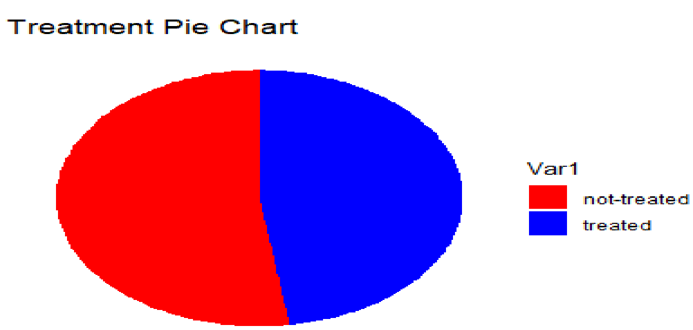
```
#Author DataFlair
data <- read.csv("C:/Users/Master Mind Computer/Desktop/R
project/Effectiveness of drugs.csv", TRUE)  #Reading CSV
table(data$treatment, data$improvement)
view(data)
#chisquare test
chisq.test(data$treatment, data$improvement, correct=FALSE)

data_table <- table(data$treatment, data$improvement)
data_table
mosaicplot(data_table, color = c("darkred", "gold"), xlab
="treatment", ylab = "improvement")
library(ggplot2)

# Bar plot for improvement
ggplot(data = as.data.frame(impr)) +
  geom_col(aes(x = Var1, y = Freq), fill = "red") +
  labs(title = "Improvement Bar Plot", x = "Improvement", y =
"Count")
# Pie chart for treatment
ggplot(data = as.data.frame(trt)) +
  geom_bar(aes(x = "", y = Freq, fill = Var1), stat = "identity") +
  scale_fill_manual(values = c("red", "blue")) +
  labs(title = "Treatment Pie Chart") +
  coord_polar(theta = "y") +
  theme_void()
```

output

```
> #Author DataFlair
> data <- read.csv("C:/Users/Master Mind Compute
r/Desktop/R project/Effectiveness of drugs.csv",
TRUE)  #Reading CSV
> table(data$treatment, data$improvement)

              improved not-improved
  not-treated       26           29
  treated           35           15
> view(data)
> chisq.test(data$treatment, data$improvement, c
orrect=FALSE)

	Pearson's Chi-squared test

data:  data$treatment and data$improvement
X-squared = 5.5569, df = 1, p-value = 0.01841

> data_table <- table(data$treatment, data$impro
vement)
> data_table

              improved not-improved
  not-treated       26           29
  treated           35           15
> mosaicplot(data_table, color = c("darkred", "g
old"), xlab ="treatment", ylab = "improvement")
```


data_table
not-treated
treated
improved
not-improved
improvement
treatment


Improvement Bar Plot
60
40
20
0
Count
improved
not-improved
Improvement


Treatment Pie Chart
Var1
not-treated
treated

RESULT:

We have a chi-squared value of 5.5569. Since we get a p-Value less than the significance level of 0.05, we reject the null hypothesis and conclude that the two variables are in fact dependent.