Assignment-1

Bidit Sadhukhan

Statistics-2 B2230022 RKMVERI

2023-06-28

Table of contents

Question-1

Suppose that we are interested in assessing the effectiveness of a Social Skills Training (SST) program for alcoholics that are in a rehabilitation program. Assume that 12 patients (the control group) participated in the normal treatment program, and 11 patients (the test group) participated in the SST supplement in addition to the normal treatment program. Based on the following data, can we conclude whether SST training programme is effective or not?

Control	1042	1617	1180	973	1552	1251	1151	1511	728	1079	951	1319
Case	874	389	612	798	1152	893	541	741	1064	862	213	

Solution

For the given question, we will do the non-parametric test to check if the SST training programme is effective or not based on the given data because we are not given the information regarding the data distribution and the sample size is also small.

U statistic:

$$U_{\text{mean}} = \frac{1}{2} \cdot n_1 \cdot n_2$$

$$u_\sigma = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{n_1}}$$

Test Statistic:

$$T = \frac{U - U_{\text{mean}}}{U_{\sigma}}$$

```r
# Control group data
control <- c(1042, 1617, 1180, 973, 1552, 1251, 1151, 1511, 728, 1079, 951, 1319)

# Case group data
case <- c(874, 389, 612, 798, 1152, 893, 541, 741, 1064, 862, 213)

# Perform Mann-Whitney U test
result <- wilcox.test(control, case)
```

Output

```r
# Print the test results
print(result)
```

```
    Wilcoxon rank sum exact test

data:  control and case
W = 117, p-value = 0.0009807
alternative hypothesis: true location shift is not equal to 0
```

Comment

Based on the obtained p-value of 0.0009807, which is less than the conventional significance level of 0.05, we can reject the null hypothesis. Therefore, we have evidence to support the conclusion that there is a significant difference in the outcome between the control and case groups. This implies that the Social Skills Training (SST) program is effective in improving social skills for alcoholics in the rehabilitation program.

Question-2

Some students of an Indian City, who were interviewed during a sample survey, are classified below according to their smoking and tea drinking habits. Calculate Yule's measures of association.

	Smoker	Non-smoker
Drinks-Tea	40	33
Does not drink tea	3	12

Solution

```r
# Create a contingency table
smoking <- c(40, 3)
non_smoking <- c(33, 12)

contingency_table <- matrix(c(smoking, non_smoking), nrow = 2, ncol = 2, byrow = F)
```

```r
colnames(contingency_table) <- c("Smoker", "Non-Smoker")
rownames(contingency_table) <- c("Drinks tea", "Does not drink tea")
print(contingency_table)
```

```
                  Smoker Non-Smoker
Drinks tea            40         33
Does not drink tea     3         12
```

```r
# Calculate Yule's measures of association
a <- contingency_table[1, 1]
b <- contingency_table[1, 2]
c <- contingency_table[2, 1]
d <- contingency_table[2, 2]

Q <- (a * d - b * c) / (a * d + b * c)
Y <- sqrt(Q)
```

Output

```r
# Print the results
cat("Yule's Q:", Q, "\n")
```

```
Yule's Q: 0.6580311
```

```r
cat("Yule's Y:", Y, "\n")
```

```
Yule's Y: 0.8111912
```

Comment

Yule's Q indicates a moderate positive association between smoking and tea drinking habits. This means that individuals who smoke are more likely to be tea drinkers, and vice versa. Yule's Y, which is the square root of Yule's Q, also shows a moderate association between smoking and tea drinking.