Shenduo Zhang

Chi-square independent test

Jackson et al. (2013) wanted to know whether it is better to give the diphtheria, tetanus and pertussis (DTaP) vaccine in either the thigh or the arm, so they collected data on severe reactions to this vaccine in children aged 3 to 6 years old. One nominal variable is severe reaction vs. no severe reaction; the other nominal variable is thigh vs. arm.

1.1 Data

	No server reaction	Severe reaction
Thigh	4758	30
Arm	8840	76

1.2 Manually computation

We want to test that if the having a severe reaction is indepent from where you inject the vaccine. Denote the probability to have a severe reaction when inject in thigh as p_1 , that of in arm as p_2. And denote the total proportion of severe reaction as p. We want to test the hypothesis $H_o := p_1 = p_2 = p$ against $H_1 := p_1 \neq p_2$.

Denote the number of samples having no severe reaction as N_1, the other as N_2. Denote the sample given shot in thigh as M_1, in arm as M_2.

For this sample we have total of $N_1 + N_2 = M_1 + M_2 = 13704$ of them, and 106 of them had reactions. So the null hypothesis is $p_1 = p_2 = p = 0.7735\%$.

The expectation table is given as follow

	Expected No server reaction	Expected Severe reaction
Thigh	4750.965	37.035
Arm	8847.035	68.9652

Hence the test statistics

$$X^2 = \sum \frac{(|O - E| - 0.5)^2}{E} = 1.7862 \tag{1}$$

And the p-value is 0.1813, which is not significant. We shall not reject the test.

1.3 Programming result

First we try what will happen if we use the Fisher-Exact test on this problem,

```
> VaccineData <- matrix(c(4758,30,8840,76), ncol = 2)
> Fisher_exact_test(VaccineData)
[1] "You better choose Chi-Square test instead"
```

We can see the program finds the expected value too large for Fisher-exact test, so recommend Chi-square test to us. Then we run the Chi-square test then find the following result

```
> Chi_square_test(VaccineData,0.05)
$`value of statistics`
[1] 1.786206

$`p-value`
[1] 0.1813891

$`Reject?`
[1] FALSE
```

From the result, we shall reject not the null hypothesis.

1.4 Biological conclusion

There is no significant difference on severe reaction rates between rejecting arms or thigh after giving shots of DTaP.

Fisher-Exact test

Van Nood et al. (2013) studied patients with Clostridium difficile infections, which cause persistent diarrhea. One nominal variable was the treatment: some patients were given the antibiotic vancomycin, and some patients were given a fecal transplant. The other nominal variable was outcome: each patient was either cured or not cured.

We want to know whether the proportion for one variable are different among values of the other variable.

2.1 Data

	fecal	vancomycin
sick	2	5
cured	13	4

2.2 Manually computation

We want to test the hypothesis that the proportion at one variable are the same for different values of the second variable. Using similar notation as the last problem, we have the null hypothesis $H_0 := p_1 = p_2$. And we want to test it agains all three alternative hypothesises $H_1 := p_1 \neq p_2, p_1 < p_2, p_1 > p_2$.

The table has been rearranged so that (1,1) is the smallest number.

Adds up the probability of having 0, 1 or 2 sick samples in fecal transplant is 0.5862. Hence the p-value for testing hypothesis H_0 against $H_1 := p_1 < p_2$, which is 0.5862.

And the probability of having $2, 3, ..., 7$ sich samples is 0.7851. This gives us the other p-value for testing H_0 against $H_1 := p_1 > p_2$, which is 0.7851.

And to test H_0 against $H_1 := p_1 \neq p_2$, the p-value is given as 1. Since the lower-tail and upper tail both exceed 0.5.

In neither cases shall we reject the null hypothesis.

2.3 Programming result

First let apply Chi-square test to it to see what happened

```
> TranplantData <- matrix(c(2,5,13,4),ncol = 2)
> Chi_square_test(TranplantData,0.05)
[1] "Approximation of Chi-square is not eligbable, exists expected value smaller that 5"
```

The program recognize the expected sample value is too small for Chi-square test to recommend us to use Fisher-exact test.

The result of testing against both three alternative are given as follow.

```
> Fisher_exact_test(TranplantData,0.05,"=")
$`two-sided p-value`
[1] 1

$`Reject?`
[1] FALSE

> Fisher_exact_test(TranplantData,0.05,">")
$`right-handed p-value`
[1] 0.5862284

$`Reject?`
[1] FALSE

> Fisher_exact_test(TranplantData,0.05,"<")
$`left-handed p-value`
[1] 0.7851051

$`Reject?`
[1] FALSE
```

Which is the same as what we saw in mannual computation.

2.4 Biological conclusion

There's no significant statistical clear difference on the amount of the cured betweeen fecal and vancomycin transplant. You can say that fescal or vancomycin transplant has no impact on the curing rate.

McNemar test, Normal-theorey test

A sample of $N = 100$ cases of identical twins in which only one twin has lung cancer is selected for further study. The twin with lung cancer is the case. The other twin serves as the control. Each pair of twins is surveyed to determine if they smoke tobacco.

3.1 Data

	Smoke	No smoke
Smoke	16	21
No smoke	4	59

3.2 Mannual computation

We want to test the hypothesis of the probability of obtaining a type A discordant pair is equal to the one of type B, agains not equal to.

First we compute the quantity n_A, n_B and then plug in to compute X^2 statistics, abtain the value of $X^2 = 10.24$. By refering to the table of Chi-square distribution, we have the p-value of 0.0014

Thus we can reject the null hypothesis that there is a no difference betweeen obtaining type A discordant and type B discordant. Obtain that given significance level 0.05, one can think smoking is associated with lung cancer.

3.3 Programming result

First we put it in the program, and the program recognize it to adapt Normal theory test. And the result is given as follow.

```
> SmokeData <- matrix(c(16,21,4,59),ncol = 2)
> Mcnemar_normal_test(SmokeData,0.05)
[[1]]
[1] "This is normal test"

$`Value of statistics X^2`
[1] 10.24

$`p-value`
[1] 0.001374276

$`reject null hypothesis?`
[1] TRUE
```

Which is the same from our mannual computation.

3.4 Biological conclusion

We can think that the smoke has a significant impact on the probability of obtain a lung cancer.

McNemar test, Exact test

Question 10C on the book

A twin design is used to study age-related macular degeneration(AMD), a common eye disease of the elderly that results in substantial losses in vision. suppoese we contact 66 twinships in which one twin has AMD and the other twin does not. The twin are given a dietary questionnaire to report their usual diet. We find that in 10 twinships the AMD twin takes multivitamin supplements and the normal twin does not. In 8 twinships the normal twin takes multivitamin supplements and the AMD twin does not.j In 3 twinhsips both twins take multivitamin supplements, and in 45 twinships neither twin takes multivitamin supplements.

4.1 Data

	take multivitamin supplements	not taking
take multivitamin supplements	8	8
not taking	3	45

4.2 Mannual computation

We want to test the hypothesis of the probability of obtaining a type A discordant pair is equal to the one of type B, agains not equal to.

First we compute the quantity n_A, n_B, then we find we shall use the exact test. Since $n_A \leq n_B$ we compute the upper tail for p-value. The exact p-value is 0.227, which is far away from significant. So we shall not reject the null hypothesis.

4.3 Programming result

First we put it in the program, and the program recognize it to adapt exact test. And the result is given as follow.

```
> Mcnemar_normal_test(AMDData,0.05)
[[1]]
[1] "This is an exact test"

$`p-value`
[1] 0.2265625

$`reject null hypothesis?`
[1] FALSE
```

Which is the same from our mannual computation.

4.4 Biological conclusion

So AMD and taking multivitamin supplements are not significant associated.