HW4

Jinhong Du, 15338039

目录

1 Hypertension

Polyunsaturated fatty acids in the diet favorably affect several risk factors for cardiovascular disease. The principal dietary polyunsaturated fat is linoleic acid. To test the effects of dietary supplementation with linoleic acid on blood pressure, 17 adults consumed 23 g/day of safflower oil, high in linoleic acid, for 4 weeks. Systolic blood pressure (SBP) measurements were taken at baseline (before ingestion of oil) and 1 month later, with the mean values over several readings at each visit given in Table 9.10.

表 1: Effect of linoleic acid on SBP

Subject	Baseline SBP	1-month SBP	Baseline - 1-month SBP
1	119.67	117.33	2.34
2	100.00	98.78	1.22
3	123.56	123.83	–0.27
4	109.89	107.67	2.22

Subject	Baseline SBP	1-month SBP	Baseline - 1-month SBP
5	96.22	95.67	0.55
6	133.33	128.89	4.44
7	115.78	113.22	2.56
8	126.39	121.56	4.83
9	122.78	126.33	–3.55
10	117.44	110.39	7.05
11	111.33	107.00	4.33
12	117.33	108.44	8.89
13	120.67	117.00	3.67
14	131.67	126.89	4.78
15	92.39	93.06	–0.67
16	134.44	126.67	7.77
17	108.67	108.67	0.0

1.1 9.19 What parametric test could be used to test for the effect of linoleic acid on SBP?

The paired t test for

$$H_0 : \Delta = 0 \qquad\qquad H_1 : \Delta \neq 0$$

where Δ is the underlying mean of d - the difference between baseline and 1-month SBP.

1.2 9.20 Perform the test in Problem 9.19, and report a p-value.

```
x <- c(119.67,100.00,123.56,109.89,96.22,133.33,
       115.78,126.39,122.78,117.44,111.33,117.33,
       120.67,131.67,92.39,134.44,108.67)
y <- c(117.33,98.78,123.83,107.67,95.67,128.89,
       113.22,121.56,126.33,110.39,107.00,108.44,
       117.00,126.89,93.06,126.67,108.67)
d <- c(2.34,1.22,-0.27,2.22,0.55,4.44,2.56,
          4.83,-3.55,7.05,4.33,8.89,3.67,
          4.78,-0.67,7.77,0.00)
t.test(x, y, paired = T)
```

```
##
##  Paired t-test
##
## data:  x and y
```

```
## t = 3.7204, df = 16, p-value = 0.001861
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.269304 4.631872
## sample estimates:
## mean of the differences
##                2.950588
```

The p-value is 0.001861.

1.3 9.21 What nonparametric test could be used to test for the effect of linoleic acid on SBP?

The Wilcoxon Signed-Rank Test.

1.4 9.22 Perform the test in Problem 9.21, and report a p-value.

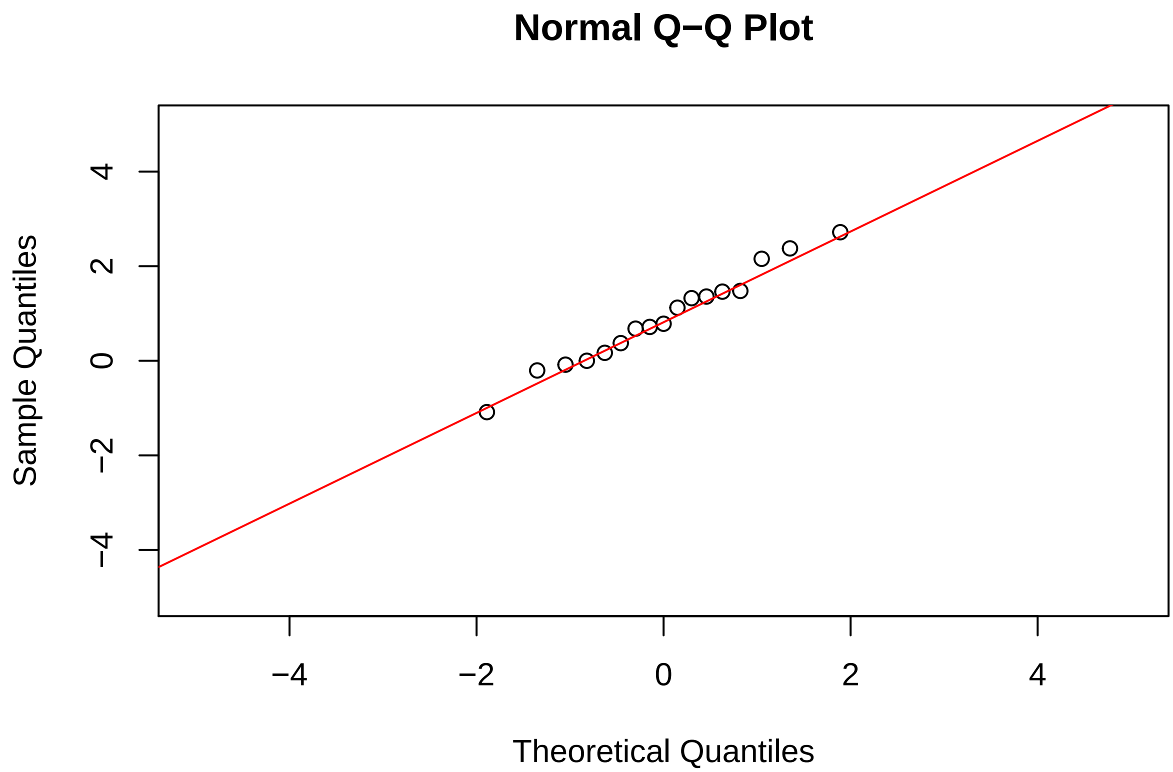
```
wilcox.test(x-y, alternative = 'two.sided', exact = F, correct = F)
```

```
##
##  Wilcoxon signed rank test
##
## data:  x - y
## V = 124, p-value = 0.003783
## alternative hypothesis: true location is not equal to 0
```

The p-value is 0.003783.

1.5 9.23 Compare your results in Problems 9.20 and 9.22, and discuss which method you feel is more appropriate here.

```
qqnorm(d/sd(d), ylim=c(-5,5), xlim = c(-5,5))
qqline(d/sd(d), col = 2)
```



The first test is more significant than the second one.

Since $d \sim N(\Delta, \sigma_d^2)$ and $d \overset{H_0}{\sim} N(0, \sigma_d^2)$ and the sample distribution of $d_1, d_2, \cdots, d_{17}$ are approximately normal distributed from QQ-plot, it is more appropriate to use the pair t test.

2 Cardiovascular Disease

An investigator wants to study the effect of cigarette smoking on the development of myocardial infarction (MI) in women. In particular, some question arises in the literature as to the relationship of timing of cigarette smoking to development of disease. One school of thought says that current smokers are at much higher risk than ex-smokers. Another school of thought that says a considerable latent period of nonsmoking is needed before the risk of ex-smokers becomes less than that of current smokers. A third school of thought is that ex-smokers may actually have a higher incidence of MI than current smokers because they may include more women with some prior cardiac symptoms (e.g., angina) than current smokers. To test this hypothesis, 2000 disease-free currently smoking women and 1000 disease-free ex-smoking women, ages 50—59, are identified in 1996, and the incidence of MI between 1996 and 1998 is noted at follow-up visits 2 years later. Investigators find that 40 currently smoking women and 10 ex-smoking women have developed the disease.

2.1 10.18 Is a one-sample or a two-sample test needed here?

It is a two-sample test since data come from two different populations - currently smoking and ex-smoking women and the purpose is to test whether the underlying parameters of two different populations are the same

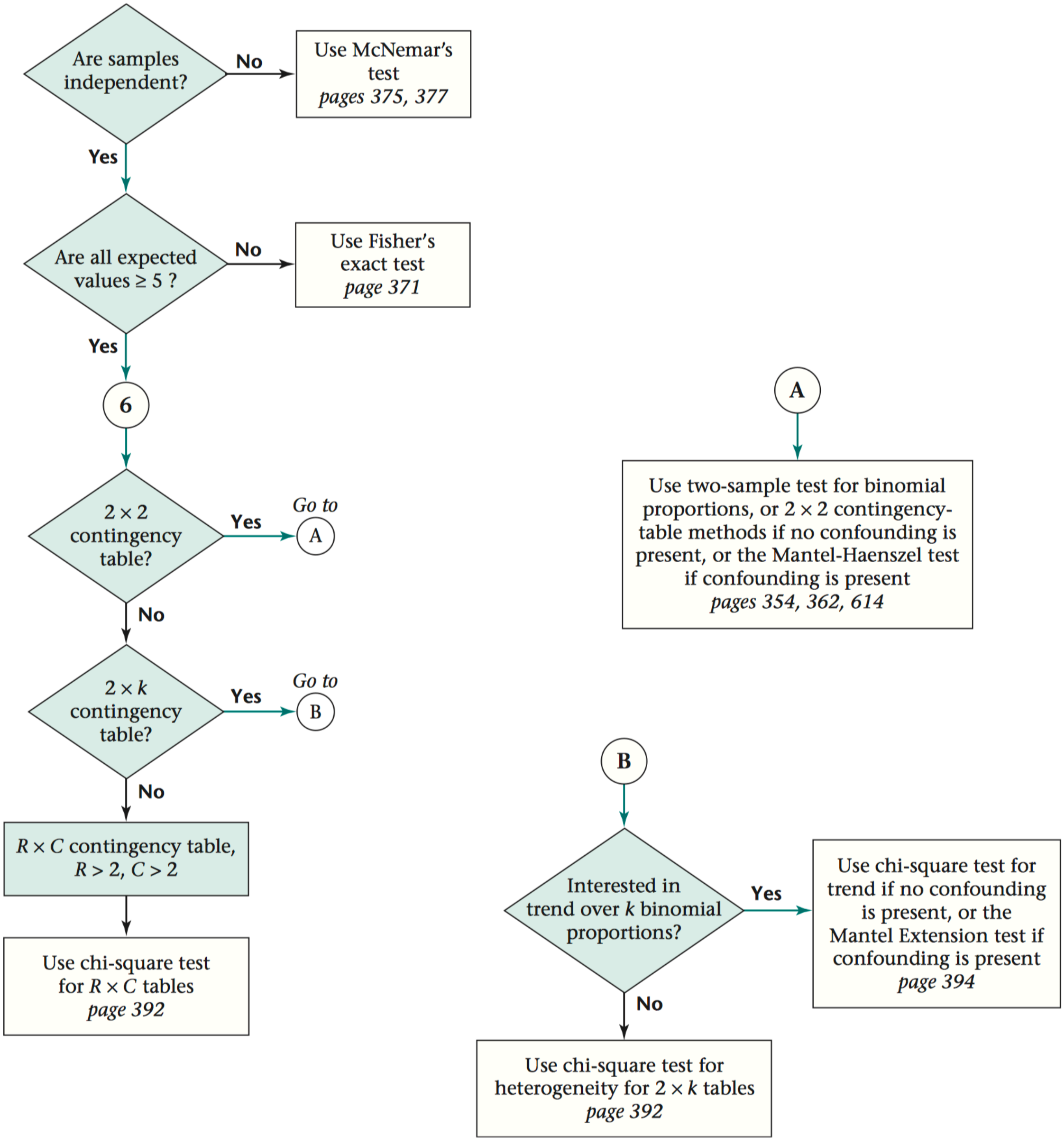


图 1: Flowchart for appropriate methods of statistical inference for categorical data or not.

表 2: Contingency Table

Status	currently smoking	ex-smoking	Total
disease	40	10	50
not disease	1960	990	2950
Total	2000	1000	3000

2.2 10.19 Is a one-sided or a two-sided test needed here?

It is a two-sided test since we want to test whether there is an effect of timing of cigarette smoking on the development of disease. More specifically, we want to test whether the incidences of MI in each groups are the same or not.

$$H_0 : p_1 = p_2 \qquad H_1 : p_1 \neq p_2$$

where p_1 and p_2 are the theoretic probability of developing diseases of currently smoking group and ex-smoking group.

2.3 10.20 Which of the following test procedures should be used to test this hypothesis? (More than one may be necessary.) (Hint: Use the flowchart in Figure 10.16.)

(1) χ^2 test for 2×2 contingency tables

(2) Fisher's exact test

(3) McNemar's test

(4) One-sample binomial test

(5) One-sample t test

(6) Two-sample t test with equal variances

Expectation values for the 2×2 contingency table is given by

```r
O <- matrix(c(40,1960,10,990),2,2)
row_margins <- apply(O,1,sum)
column_margins <- apply(O,2,sum)
grand_total <- sum(O)
E <- tcrossprod(row_margins, column_margins) / grand_total
E
```

```
##             [,1]      [,2]
## [1,]   33.33333  16.66667
## [2,] 1966.66667 983.33333
```

Since the samples are independent identically distributed (the event that everyone develops ML or not will not be influenced by others) and the least expected number among cells in the 2×2 contingency table is bigger than 5, we go to A in the flowchart. For the above test procedures, we can only use the Yates-Corrected χ^2 test for 2×2 contingency tables.

2.4 10.21 Carry out the test procedure(s) mentioned in Problem 10.20, and report a p-value.

```r
chisq.test(O, correct = T)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  O
## X-squared = 3.4805, df = 1, p-value = 0.0621
```

Since p-value is bigger than 0.05, so with significance levle $\alpha = 0.05$, we accept H_0, i.e., there is no difference between currently smoking women and ex-smoke women in this study.