STAT 3004: Solutions of Assignment 2

Cheng Li
16 Oct 2020

---

1 Problem 1

From the problem, we can use one-sample t-test as the parametric test and Wilcoxon signed-rank test as non-parametric test. So let $d_i = x_i - y_i$ and δ be the median of $\{d_i\}$, where x_i is the i-th BMD for femoral neck of ligher-smoking twin and y_i is the i-th BMD for fermoral neck of heavier- smoking twin. Therefore, we can get two hypothesis:

$$H_0 : \mu = 0,$$
$$H_1 : \mu \neq 0,$$

and

$$H_0 : \delta = 0,$$
$$H_1 : \delta \neq 0,$$

where μ means the mean of d_i. With R, we can easily get p-value of one-sample t-test is $0.96 > 0.05$ and p-value of Wilcoxon signed-rank is $0.856 > 0$. Thus, we can not reject either of null-hypothesis.

2 Problem 2

2.1 (a)

We can use the sign test. It is easily for us to get that the test statistics C is 15 and n is 23. So, we use the normal theory test. The rejection region is given by $C > c_{upper}$ or $C < clower$, where

$$\begin{aligned} c_{upper} &= \frac{n}{2} + \frac{1}{2} + z_{0.975}\sqrt{\frac{n}{4}} \\ &= 16.7, \\ c_{lower} &= \frac{n}{2} - \frac{1}{2} - z_{0.975}\sqrt{\frac{n}{4}} \\ &= 6.3. \end{aligned}$$

Therefore, we can not reject H_0 at the 5% level. But we assumed that the periodontal status of patients would remain unchanged in the absence of the program, which is a questionable assumption. A better study design would involve following a control group over 6 months who did not receive the education program and comparing results in the two groups.

2.2 (b)

From the question, we can get the sum rank R of d^+ is 185. So, the test statistics is given by

$$T = \frac{\left|R - \frac{n(n+1)}{4}\right| - 0.5}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \sum_{i=1}^{g} \frac{(t_i^3 - t_i)}{48}}}$$
$$= 1.436 \sim N(0,1) \; under \; H_0$$

So the p-value is 0.151. Thus, the periodontal status of the patients has not significantly changed over time, even when accounting for the magnitude of improvement or decline.

2.3 (c)

The normal theory test can be used, since $min(n_1, n_2) = 12 \leq 10$. The test statistic is given by

$$T = \frac{\left|R - \frac{n_1(n_1+n_2+1)}{2}\right| - 0.5}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$
$$= 2.513 \sim N(0,1) \; under \; H_0$$

Thus, the p-value is 0.012 less than 0.05. So, we should reject H_0.

3 Problem 3

3.1 (a)

We should use the chi-square test. and the hypothesis is given by

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

where

$$p_1 = Prob(otorrhea) \; for \; the \; ear \; drop \; group,$$
$$p_2 = Prob(otorrhea) \; for \; the \; observation \; group.$$

3.2 (b)

We can form the 2*2 table relating outcome to group as follows:

Group	Otorrhea 2 weeks = yes	Otorrhea 2 weeks = no	Total
Ear drop	4	72	76
Observation	41	34	75
Total	45	106	151

Table 1: Observed Table

The expected counts under the null hypothesis are as follows:

$$E_{11} = 22.65,$$
$$E_{12} = 53.35,$$
$$E_{21} = 22.35,$$
$$E_{22} = 52.65.$$

Thus, the expected table is as follows:

Group	Otorrhea 2 weeks = yes	Otorrhea 2 weeks = no	Total
Ear drop	22.65	53.35	76
Observation	22.35	52.65	75
Total	45	106	151

Table 2: Expected Table

Since all expected counts are not less than 5, we can use the chi-square test for 2*2 tables. So the statistic is given by

$$\begin{aligned}\chi^2_{corr} &= \frac{(|4-22.65|-0.5)^2}{22.65} + \frac{(|72-53.35|-0.5)^2}{53.35} + \frac{(|41-22.35|-0.5)^2}{22.35} + \frac{(|34-52.65|-0.5)^2}{52.65} \\ &= 41.71 \sim \chi^2_1 \; under \; H_0.\end{aligned}$$

Since $\chi^2_{1,0.999} = 10.83 < 41.71$, which means that $p < 0.001$. Thus, there is a highly significant difference in prevalence between the 2 groups.

4 Problem 4

4.1 (a)

We can form a 2*2 table relating the type of bird to the type of sunflower seeds eaten:

	Type of seed		
Type of Bird	black oil	striped	total
Titmouse	1	4	5
Gold Finch	19	5	24
Total	20	9	29

Table 3: Observed Table

The smallest expected value in this table is $E_{12} = 5 * \frac{9}{29} = 1.55 < 5$.Thus, we should use Fisher's exact test to test the hypothesis:

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

where p_1 is the proportion of titmice who prefer black oil seeds and p_2 is the proportion of gold finches who prefer black oil seeds.

4.2 (b)

To perform this test, we need to enumerate all possible tables with the same row and column margins as the observed table:

0	5
20	4
1	4
19	5
2	3
18	6
3	2
17	7
4	1
16	8
5	0
15	9

We can get $Pr(0) = 0.001$,$Pr(1) = 0.021$,$Pr(2) = 0.134$,$Pr(3) = 0.346$,$Pr(4) = 0.367$,$Pr(5) = 0.131$. Since the observed table is the ”1” table, the two-tailed p-value = $2*(0.001+0.021) = 0.045$.Therefore, we should reject H_0 at the level 0.05.

4.3 (c)

We display the observed and expected counts in a 2*4 table as shown below(expected counts in parentheses):

		Day				
		1	2	3	4	Total
Type of Seed	black oil	19 (14.2)	14 (14.2)	9 (8.88)	45 (49.71)	87
	striped	5 (9.8)	10 (9.8)	6 (6.12)	39 (34.29)	60
	Total	24	24	15	84	147

The smallest expected value is 6.1 ¿ 5. Thus, we can use the chi-square test for R*C tables to test the hypothesis

$$H_0 : p_1 = p_2 = p_3 = p_4,$$
$$H_1 : at\ least\ two\ of\ the\ p_i\ are\ different,$$

where p_is are proportion of gold finches who prefer black oil seeds on the ith day, $i = 1, ..., 4$.

4.4 (d)

The expected value for the E_{ij} cell(listed in parentheses in the above table) is obtained from $E_{ij} = R_iC_j/N, i = 1, 2; j = 1, 2, 3, 4$, where R_i is i-th row total, C_j is j-th column total. So we have the test statistic $\chi^2 = 5.07 \sim \chi^2_3\ under\ H_0$. Since 5.07 is less than $\chi^2_{3,0.95}$, p-value is larger than 0.05. Thus, there is no significant difference in feeding preferences by day.