2.1

maksimov.denis.o

May 2020

1 Задача 2.1

Выборка:

Рассматривается задача тестирования вакцины от некоторого вируса. Производство вакцины очень дорогое и затратное по времени, поэтому в день может быть произведена только одна ампула.

Требуется проверить, что вакцина помогает (вероятность заразиться меньше у человека с вакциной чем у человека без вакцины).

Эксперимент ставится следующим образом: каждый день есть два идентичных по здоровью человека. Один из людей принимает вакцину, а второй нет, после чего обоих ставят в одну среду с вирусом. В конце проверяют, кто заразился. (В таблице: s — sick; h — healthy)

Весь мир ждет вакцину от данного вируса, поэтому к руководству института постоянно приходят запросы о сроках завершения тестирования образца. Руководство поручило Вам оценить среднее время, которое понадобится на тестирования данной вакцины. А также провести анализ полученных данных на уровне значимости $\alpha = 0.05$ и при ошибке второго рода $\beta = 0.8$.

Требуется:

- Записать задачу формально;
- Выполнить оценку среднего количества дней для принятия решения (учесть что истинная вероятность заразиться с вакциной и без равны $p_1 = 0.2$, $p_2 = 0.5$ соответственно);
- Выполнить анализ данных и выяснить работает ли вакцина или нет.

Все выкладки должны быть сделаны аналитически, без использования компьютера.

2 Задача формально

Имеем последовательное сэмплирование парами из двух связанных бернуллиевских выборок.

Изначально:
нулевая гипотеза: $H_0 \ : \ p_1 \geq p_2$
альтернатива: $H_1 \ : \ p_1 < p_2$
В терминах последовательного анализа, выбрав пороги:
нулевая гипотеза: $H_0 \ : \ u \geq u_U = 2$
альтернатива: $H_1 \ : \ u \leq u_L = \frac{1}{2}$

3 Решение

Провернем последовательный аналог Z-критерия для разности двух долей (связанные выборки).

Обозначим:

$$A = \frac{1-\beta}{\alpha} = \frac{1-0.8}{0.05} = 4, \qquad B = \frac{\beta}{1-\alpha} = \frac{0.8}{0.95} = \frac{16}{19}$$

$$u = \frac{p_1(1-p_2)}{p_2(1-p_1)} = \frac{0.2(1-0.5)}{0.5(1-0.2)} = \frac{1}{4}$$

$$\mathbb{E}_u(n) = \frac{L(u)\ln B + (1-L(u))\ln A}{\left(\frac{u}{u+1}\ln\frac{u_U(1+u_L)}{u_L(1+u_U)} + \frac{1}{u+1}\ln\frac{1+u_L}{1+u_U}\right)(p_1(1-p_2)+p_2(1-p_1))}$$

$$L(u) = \frac{A^h - 1}{A^h - B^h}$$

$$\frac{u}{u+1} = \frac{1-\left(\frac{1+u_L}{1+u_U}\right)^h}{\left(\frac{u_U(1+u_L)}{u_L(1+u_U)}\right)^h - \left(\frac{1+u_L}{1+u_U}\right)^h} = \frac{\left(\frac{1+u_U}{1+u_L}\right)^h - 1}{\left(\frac{u_L}{u_U}\right)^h - 1} \to \frac{1}{5} = \frac{2^h-1}{4^h-1} = \frac{1}{2^h+1} \to h = 2$$

$$L(u) = \frac{4^2-1}{4^2 - \left(\frac{16}{19}\right)^2} = \frac{(4-1)(4+1)}{4^2(1-\frac{4}{19})(1+\frac{4}{19})} = \frac{19^2}{23*4^2} = \frac{361}{368}$$

$$\mathbb{E}_u(n) = \frac{\frac{361}{368}(4\ln 2 - \ln 19) + \frac{7}{184}\ln 2}{-0.3\ln 2} = \frac{\frac{361}{368}\ln 19 - \frac{729}{184}\ln 2}{0.3\ln 2} = \frac{5}{552}\Big(361\log_2 19 - 1458\Big) \approx 0.68$$

Результат странный, скорее всего из-за большого β : будем счиать, что $1-\beta = 0.8$

$$A = \frac{1-\beta}{\alpha} = \frac{0.8}{0.05} = 16, \qquad B = \frac{\beta}{1-\alpha} = \frac{0.2}{0.95} = \frac{4}{19}$$

$$L(u) = \frac{16^2-1}{16^2 - \left(\frac{4}{19}\right)^2} = \frac{(16-1)(16+1)}{4^2(4-\frac{1}{19})(4+\frac{1}{19})} = \frac{17*19^2}{77*5*4^2} = \frac{6137}{6160} \approx 0.996$$

$$\mathbb{E}_u(n) = \frac{\frac{6137}{6160}(2\ln 2 - \ln 19) + \frac{23}{6160}4\ln 2}{-0.3\ln 2} = \frac{1}{616*3}\frac{6137\ln 19 - 12366\ln 2}{\ln 2} = \frac{1}{1848}\Big(6137\log_2 19 - 12366\Big) \approx 7.4$$

Т.е. в среднем понадобится 7.4 дней

$$a_m = \frac{\ln B + m\ln\frac{1+u_U}{1+u_L}}{\ln u_U - \ln u_L} = \frac{2\ln 2 - \ln 19 + m\ln 2}{2\ln 2} = 1 - \frac{1}{2}\log_2 19 + \frac{m}{2} \approx -1.124 + \frac{m}{2}$$

$$r_m = \frac{\ln A + m\ln\frac{1+u_U}{1+u_L}}{\ln u_U - \ln u_L} = \frac{4\ln 2 + m\ln 2}{2\ln 2} = 2 + \frac{m}{2}$$

$$d_m = \sum_{i=1}^{m}(1-X_{1i})X_{2i}$$

- Если $d_m \geq r_m$, то отвергаем H_0;
- если $d_m \leq a_m$, то принимаем H_0;
- иначе - продолжаем процесс

[49]:
```
import pandas as pd
```

[68]:
```
data = pd.read_csv("2.1.csv")
data
```

[68]:

	with vaccine	without vaccine
0	h	h
1	h	s
2	h	s
3	h	h

4	h	h
5	h	s
6	h	h
7	h	s
8	h	s
9	h	s
10	s	s
11	h	h
12	h	s
13	h	h
14	h	h
15	h	s
16	h	h
17	s	h
18	h	s
19	h	h
20	h	s
21	h	h
22	h	h
23	h	h
24	s	s
25	h	s
26	s	s
27	h	s
28	h	h
29	s	s

$$h = 0, s = 1$$

Будем пропускать те дни, когда нет разницы

$$d_1 = 1, \quad a_1 = -0.624, \quad r_1 = 2.5$$

$$d_2 = 2, \quad a_2 = -0.124, \quad r_2 = 3$$

$$d_3 = 3, \quad a_3 = 0.376, \quad r_3 = 3.5$$

$$d_4 = 4, \quad a_4 = 0.876, \quad r_4 = 4$$

Четвертое несовпадение состоялось в 8 день.

4 Вывод:

Таким образом, удалось за восемь дней испытаний показать, что вакцина действует. Это согласуется со прогнозом $\mathbb{E}_u(n) = 7.4$.