

Задача 2.1

Северилов Павел, 674

15 мая 2020 г.

1 Задача

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 дня 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 $\beta = 0.2$. Вероятность заразиться с вакциной и без равны $p_1 = 0.2, p_2 = 0.5$ соответственно.

2 Решение

2.1 Формальная постановка задачи

Наблюдения поступают группами по 2 элемента.

нулевая гипотеза: $H_0 : u \geq u_U$

альтернатива: $H_1 : u \leq u_L$

статистика: $d_m(X_1^m, X_2^m) = \sum_{i=1}^m (1 - X_{1i}) X_{2i}$

X_{1i}, X_{2i} – значения, поступающие парами. $k_1 = \frac{p_1}{1-p_1}, k_2 = \frac{p_2}{1-p_2}$ – риски, $u = \frac{k_1}{k_2} = \frac{p_1(1-p_2)}{p_2(1-p_1)}$ – относительный риск:

- $u = 1 \Leftrightarrow p_1 = p_2$
- $u > 1 \Leftrightarrow p_1 > p_2$
- $u < 1 \Leftrightarrow p_1 < p_2$

Фиксируем "коридор" отклонений u от 1, которые можно считать незначимыми: $u_L \leq 1 \leq u_U$

2.2 Выкладки

Имеем $\alpha = 0.05, \beta = 0.2$. Тогда:

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

Также

$$u = \frac{p_1(1 - p_2)}{p_2(1 - p_1)} = \frac{0.2 \cdot 0.5}{0.5 \cdot 0.8} = \frac{1}{4}, \quad p_1(1 - p_2) + p_2(1 - p_1) = 0.5$$

Момент остановки:

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

где $L(u) = \frac{A^h - 1}{A^h - B^h}$, а h находим из

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

Упростим:

$$\frac{1/4}{5/4} = \frac{\left(\frac{1+u_U}{1+u_L}\right)^h - 1}{\left(\frac{u_U}{u_L}\right)^h - 1} \Rightarrow \left(\frac{u_U}{u_L}\right)^h - 1 = 5 \cdot \left(\frac{1+u_U}{1+u_L}\right)^h - 5 \Rightarrow 5 \left(\frac{1+u_U}{1+u_L}\right)^h - \left(\frac{u_U}{u_L}\right)^h = 4$$

Видно, что подходит $h = 0$, но есть и второй корень. Подберем коридор u_L, u_U удобным образом, чтобы h получился целым. Возьмем $u_L = 0.5, u_U = 2$. Тогда получим:

$$5 \cdot 2^h - 4^h = 4 \Rightarrow h = 2$$

Функция $y_1 = 5 \cdot 2^h - 4^h$ пересекает $y_2 = 4$ в двух точках, поэтому других корней h нет. Тогда

$$L(u)_{h=0} = 0, \quad L(u)_{h=2} = \frac{16^2 - 1}{16^2 - \frac{4}{19}^2} = \frac{6137}{6160} \approx 1$$

В итоге матожидание дней для остановки:

$$\mathbb{E}_u(n) = 2 \cdot \frac{L(u) \ln \left(\frac{4}{19}\right) + (1 - L(u)) \ln 16}{0.2 \ln \frac{2(1+0.5)}{0.5(1+2)} + 0.8 \ln \frac{1+0.5}{1+2}} = \frac{L(u) \ln \left(\frac{4}{19}\right) + (1 - L(u)) \ln 16}{-0.3 \ln 2}$$

При $h = 0$: $\mathbb{E}_u(n) \approx -13.3$, что не подходит по смыслу. При $h = 2$:

$$\mathbb{E}_u(n) \approx \frac{\ln \frac{4}{19}}{-0.3 \ln 2} \approx 7.49$$

Константы последовательного анализа:

$$a_m = \frac{\ln B + m \ln \frac{1+u_U}{1+u_L}}{\ln u_U - \ln u_L} = \frac{\ln \frac{4}{19} + m \ln \frac{1+2}{1+0.5}}{\ln 2 - \ln \frac{1}{2}} = \frac{m}{2} + \frac{\ln \frac{4}{19}}{2 \ln 2},$$

$$r_m = \frac{\ln A + m \ln \frac{1+u_U}{1+u_L}}{\ln u_U - \ln u_L} = \frac{\ln 16 + m \ln 2}{2 \ln 2} = \frac{m}{2} + 2$$

- $d_m \geq r_m \Rightarrow$ отвергаем H_0
- $d_m \leq a_m \Rightarrow$ принимаем H_0
- $a_m < d_m < r_m \Rightarrow$ процесс продолжается, добавляем элемент выборки.

Проведем данный процесс на данной в задаче выборке. Будем считать, что health $\leftrightarrow 0$, sick $\leftrightarrow 1$. Будем рассматривать только различающиеся пары – (0, 1) и (1, 0), а остальные будем отбрасывать.

m	d_m	номер дня	a_m	r_m
1	1	2	-0.624	2.5
2	2	3	-0.124	3.0
3	3	6	0.376	3.5
4	4	8	0.876	4.0
5	5	9	1.376	4.5
6	6	10	1.876	5.0
7	7	13	2.376	5.5

Таблица 1: Результаты

with vaccine	without vaccine
0	0
0	1
0	1
0	0
0	0
0	1
0	0
0	1
0	1
0	1

Таблица 2: Первые 10 наблюдений из выборки

Как видим, при $m = 4$ происходит первое пересечение с границей (с r_m). Это соответствует восьмому дню из выборки, т.е. принимаем решение отвергнуть нулевую гипотезу через 8 дней. Построим для наглядности по полученным данным график (m соответствует номеру наблюдений из выборки с различающимися парами, номер соответствующего дня приведен в таблице)

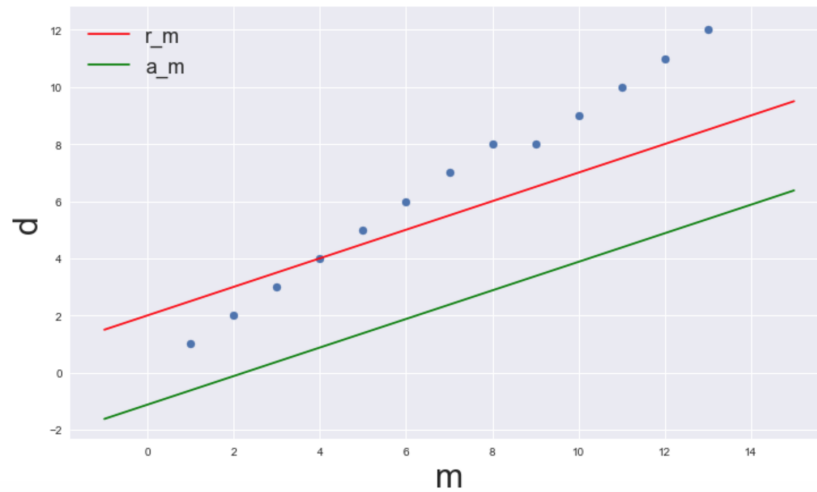


Рис. 1: asd

3 Вывод

Посчитанное аналитически матожидание дней согласуется с реальностью – 7,5 дней при реальных 8 днях. По выборке через 8 дней приняли решение отвергнуть гипотезу $u \geq u_L$, что соответствует тому, что вакцина действует и лучше плацебо.