15 мая 2020 г. Андреев Павел, 626 группа

Задача 2.1

1 Формальная постановка

$$\begin{array}{r|l} \text{Выборки:} & X_1^N = (X_{1,1},..,X_{1,N}),\ X_1 \sim Ber(p_1) \\ & X_2^N = (X_{2,1},..,X_{2,N}),\ X_2 \sim Ber(p_2) \\ & \text{выборки связанны} \\ \text{Гипотеза:} & H_0 : u = \frac{p_1(1-p_2)}{p_2(1-p_1)} \geq u_U \\ \text{Альтернатива:} & H_1 : u \leq u_L \\ \text{Статистика:} & d_m(X_1^N, X_2^N) = \sum_{i=1}^N (1 - X_{1,i}) X_{2,i} \end{array}$$

В данном случае X_1 соответствует попыткам заражения вакцинированного человека, а X_2 – невакцинированного. u_U считаем равным 2, u_L – 0.5.

2 Среднее количество дней.

Расчитаем среднее количество дней необходимое для принятия решения об эффективности вакцины (на уровне значимости $\alpha = 0.05$ и при ошибке второго рода $\beta = 0.2$). Для начала найдем параметр h из уравнения:

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

При этом учтем, что истинная вероятность заразиться с вакциной и без равны $p_1 = 0.2$, $p_2 = 0.5$ соответственно. То есть $u = \frac{0.2 \cdot 0.5}{0.8 \cdot 0.5} = \frac{1}{4}$

$$\frac{1/4}{1/4+1} = \frac{1 - \left(\dfrac{1+1/2}{1+2}\right)^h}{\left(\dfrac{2(1+1/2)}{1/2(1+2)}\right)^h + \left(\dfrac{1+1/2}{1+2}\right)^h}$$

$$\frac{1}{5} = \frac{1 - \dfrac{1}{2^h}}{4^h \cdot \dfrac{1}{2^h} - \dfrac{1}{2^h}} = \frac{2^h - 1}{4^h - 1} = \frac{1}{1 + 2^h}$$

$$h = 2$$

Матожидание числа шагов равно:

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

где $A = \frac{1-\beta}{\alpha} = \frac{1-0.2}{0.05} = 16$, $B = \frac{\beta}{1-\alpha} = \frac{0.2}{0.95} = \frac{4}{19}$, $L(u) = \frac{A^h - 1}{A^h - B^h} = \frac{16^2-1}{16^2 - 4^2/19^2} = \frac{51 \cdot 19^2}{16 \cdot 1155} \cong 0.996$

$$\mathbb{E}_u(n) \cong \frac{-1.552 + 0.011}{-0.416} \Bigg/ \frac{1}{2} \cong 7.4$$

3 Построение процедуры.

Посчитаем константы последовательного анализа:

$$a_m = \frac{\log B + m \log \dfrac{1+u_U}{1+u_L}}{\log u_U - \log u_L} = \frac{-1.558 + m \cdot 0.693}{1.386} = -1.124 + m \cdot 0.5$$

$$r_m = \frac{\log A + m \log \dfrac{1+u_U}{1+u_L}}{\log u_U - \log u_L} = \frac{2.773 + m \cdot 0.693}{1.386} = 2 + m \cdot 0.5$$

Проведем процедуру последовательного анализа имеющихся данных:

Номер дня	С вакциной	Без вакцины	a_m	r_m	d_m
1	здоров	здоров	0	-1.124	2
2	здоров	болен	1	-0.624	2.5
3	здоров	болен	2	-0.124	3
4	здоров	здоров	2	-0.124	3
5	здоров	здоров	2	-0.124	3
6	здоров	болен	3	0.376	3.5
7	здоров	здоров	3	0.376	3.5
8	здоров	болен	4	0.876	4

Как видно, на восьмом дне (что, кстати, согласуется со значением матожидания) $a_m = d_m$, то есть статистика достигла верхней границы, и мы можем отвергнуть гипотезу о неэффективности препарата в пользу альтернативы – его эффективности.