

## РАСЧЁТ КРИТЕРИЯ СОГЛАСИЯ ПИРСОНА СРЕДСТВАМИ ЯЗЫКА R

Данная статья посвящена Расчёту критерия согласия Пирсона средствами языка R. Проанализированы характерные особенности языка R, позволяющие рассчитать критерий согласия. Авторы сделали вывод о скорости расчётов средствами языка R.

### ВВЕДЕНИЕ

Так как язык R это язык программирования для статистической обработки данных, он широко используется как статистическое программное обеспечение для анализа данных. Мы решили проанализировать скорость выполнения задачи расчёта критерия согласия Пирсона средствами языка R и вручную. Для этого мы рассчитали критерий согласия через R и вручную, после чего сравнили затраченное время по выполнению задачи.

#### I. ПОСТАНОВКА ЗАДАЧИ

Представим, что мы выполнили эксперимент по установлению эффективности подавления развития микробного заболевания при введении в организм соответствующих антител. Всего в эксперименте было задействовано 111 мышей, которых мы разделили на две группы, включающие 57 и 54 животных соответственно. Первой группе мышей сделали инъекции патогенных бактерий с последующим введением сыворотки крови, содержащей антитела против этих бактерий. Животным из второй группы ввели только бактериальные инъекции. После некоторого времени инкубации оказалось, что 38 мышей погибли, а 73 выжили. Из погибших 13 принадлежали первой группе, а 25 – ко второй. Проверяемую в этом эксперименте гипотезу можно сформулировать так: введение сыворотки с антителами не оказывает никакого влияния на выживаемость мышей.

Таблица 1 – Данные из эксперимента

Группа	Погибло	Выжило	Всего
Бактерии + сыворотка	13	44	57
Только бактерии	25	29	54
Всего	38	73	111

#### II. РЕШЕНИЕ ЗАДАЧИ ВРУЧНУЮ

Для проверки сформулированной выше нулевой гипотезы нам необходимо знать, какова была бы ситуация, если бы антитела действительно не оказывали никакого действия на выживаемость мышей. Другими словами, нужно рассчитать ожидаемые частоты для соответствующих ячеек таблицы сопряженности.

Таблица 2 – Таблица сопряженности с ожидаемыми частотами

Группа	Погибло	Выжило	Всего
Бактерии + сыворотка	19.5	37.5	57
Только бактерии	18.5	35.5	54
Всего	38	73	111

Как видим, ожидаемые частоты довольно сильно отличаются от наблюдаемых, т.е. введение антител, похоже, все-таки оказывает влияние на выживаемость мышей, зараженных патогенным микроорганизмом. Это наблюдение мы можем выразить количественно при помощи критерия согласия Пирсона  $\chi^2$ :

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

где  $f_0$  и  $f_e$  - наблюдаемые и ожидаемые частоты соответственно. Суммирование производится по всем ячейкам таблицы. Так, для рассматриваемого примера имеем:

$$\chi^2 = 6.79$$

Далее необходимо найти соответствующее критическое значение критерия. Число степеней свободы для  $\chi^2$  рассчитывается как  $df = (R - 1)(C - 1)$ , где R и C - количество строк и столбцов в таблице сопряженности. В нашем случае  $df = (2 - 1)(2 - 1) = 1$  Зная число степеней свободы, мы теперь легко можем узнать критическое значение  $\chi^2$  при помощи стандартной R-функции `qchisq()`:

```
qchisq(p = 0.95, df = 1)
[1] 3.841459
```

Таким образом, при одной степени свободы только в 5% случаев величина критерия  $\chi^2$  превышает 3.841. Полученное нами значение 6.79 значительно превышает это критического значение, что дает нам право отвергнуть нулевую гипотезу об отсутствии связи между введением антител и выживаемостью зараженных мышей. Отвергая эту гипотезу, мы рискуем ошибиться с вероятностью менее 5%.

Следует отметить, что приведенная выше формула для критерия  $\chi^2$  дает несколько завь-

шенные значения при работе с таблицами сопряженности размером 2x2. Причина заключается в том, что распределение самого критерия  $\chi^2$  является непрерывным, тогда как частоты бинарных признаков ("погибло"/ "выжило") по определению дискретны. В связи с этим при расчете критерия принято вводить поправку на непрерывность, или поправку Йетса:

$$\chi_y^2 = \sum \frac{(|f_0 - f_e| - 0.5)^2}{f_e}$$

В нашем случае критерий  $\chi^2$  с поправкой на непрерывность составил бы 5.792, и нулевая гипотеза об отсутствии эффекта антител была бы отклонена.

### III. РЕШЕНИЕ ЗАДАЧИ СРЕДСТВАМИ ЯЗЫКА R

Все те вычисления, которые мы провели вручную, в языке R можно выполнить при помощи стандартной функции `chisq.test()`. При работе с этой функцией данные оформляются в виде матрицы, напоминающей приведенную выше таблицу сопряженности:

*Ермакович Максим Александрович*, студент 2 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, [tsumeshka@gmail.com](mailto:tsumeshka@gmail.com).

*Савич Станислав Игоревич*, студент 2 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, [savich.stanislav@mail.ru](mailto:savich.stanislav@mail.ru).

*Научный руководитель: Гуринович Алевтина Борисовна*, доцент кафедры вычислительных методов и программирования Белорусского государственного университета, кандидат технических наук, заместитель декана факультета информационных технологий и управления, [gurinovich@bsuir.by](mailto:gurinovich@bsuir.by).

```
> mice <- matrix(c(13, 44, 25, 29), nrow = 2, byrow = TRUE)
> mice
      [,1] [,2]
[1,]   13   44
[2,]   25   29
> chisq.test(mice)

Pearson's Chi-squared test with Yates' continuity correction

data:  mice
X-squared = 5.7923, df = 1, p-value = 0.0161
```

R автоматически применяет поправку Йетса на непрерывность. Рассчитанное программой значение  $\chi^2$  составило 5.79213. Мы можем отклонить нулевую гипотезу об отсутствии эффекта антител, рискуя ошибиться с вероятностью чуть более 1% (p-value = 0.0161).

### Список литературы

1. R: Анализ и визуализация данных [Электронный ресурс] / Сергей Мاستицкий© – 2011-2017. – Режим доступа: <https://r-analytics.blogspot.com.by/> – // Дата доступа: 20.04.2018.
2. Prognoz [Электронный ресурс] / Критерий согласия Пирсона. – АО «ПРОГНОЗ», 2018. – Режим доступа: <http://help.prognoz.com/> – Дата доступа: 20.04.2018.
3. Научная библиотека [Электронный ресурс] / Критерий «хи квадрат» Пирсона. –2018. – Режим доступа: <http://stu.alnam.ru/>