УДК 519.233.32

***А. Д. Черемухин***

*ст. преподаватель кафедры «Физико-математические науки»*

*ГБОУ ВО НГИЭУ*

**эФФЕКТИВНОСТь РАЗЛИЧНЫХ СТАТИСТИЧЕСКИХ ТЕСТОВ при проверке гипотезы равенства групповых средних при неравной дисперсии с исполЬЗОВАНИЕМ ЯЗЫКА r**

**Аннотация:** В работе проведен частичный анализ мощности статистических тестов при проверке гипотезы равенства групповых средних при условии нормальности распределения данных и неравенства групповых дисперсий. Установлено, что на малых объемах выборки наиболее эффективными являются Revised generalized F-test, One Stage Range test, B-square test

**Ключевые слова:** статистическая гипотеза, равенство средних, неравная дисперсия, ошибка первого рода, ошибка второго рода, эффективность критериев

Несмотря на активное развитие методов машинного обучения и обработки данных, ответы на некоторые классические статистические вопросы в окончательном виде не сформулированы до сих пор.

К числу таких вопросов относится и вопрос об оптимальном статистическом критерии проверки гипотезы о равенстве средних групповых значений при предположении о неравенстве дисперсии в группах и условии нормального распределения данных.

Для решения данной задачи в среде R был создан пакет doex, в котором реализовано 20 основных статистических тестов [1], позволяющих проверить данную гипотезу: Approximate F-test, Alexandern-Govern test, B-square test, Brown-Forsythe test, Box F-test, Cochran F-test, Fiducial Approach test, Generalized F-test, Johansen F-test,

Modified Brown-Forsythe test, Modified Generalized F-test, Modified Welch Test, One Stage test, One Stage Range test, Parametric Bootstrap test, Permutation F-test, Revised generalized F-test, Scott-Smith Test, Welch-Aspin test, Welch F-test.

В этой же работе было отмечено, что в результате симуляций по методу Монте-Карло для случая 3,5,7 групп на больших объемах данных наименьшую ошибку 1 типа дает Approximate F-test и модифицированный тест Брауна-Форсайта.

В нашей работе точность данных тестов оценивается по следующей методике:

1. При заданных средних значениях (10 и 11) и заданных среднеквадратичных отклонениях (1 и 2) генерируется три выборки. Первая – со средним значением 10 и среднеквадратичном отклонении 1, вторая – со средним значением 10 и среднеквадратичном отклонении 2, третья – со средним значением 11 и среднеквадратичном отклонении 2.
2. Далее с использованием выбранного теста по объединенным данным первой и третьей выборки проверяется гипотеза о равенстве средних на уровне значимости в 0.01. Если полученное p-значение критерия больше 0.01, то гипотеза отклоняется верно, если меньше 0.01 – то гипотеза ошибочно принята. По объединенным данным первой и второй выборки также проверяется данная гипотеза. Если полученное p-значение критерия меньше 0.01, то гипотеза принимается верно, если больше 0.01 – то гипотеза ошибочно отклонена.

3. После этого данная симуляция проводится 1000 раз по каждому тесту, и подводится количество разных исходов.

На основании полученных результатов были рассчитаны следующие метрики качества (по аналогии с показателями качества решенной задачи классификации):

- accuracy (общая точность тестов, процент правильно обработанных выборок);

- precision (специфичность, процент правильно принятых гипотез)

- recall (чувствительность, способность теста правильно принимать гипотезу);

- F-мера (гармоническое среднее precision и recall, общая характеристика точности качества).

Результаты проведенного исследования представлены в таблице 1.

Таблица 1 – Рассчитанные показатели точности качества проверки гипотез о равенстве средних с помощью тестов из пакета doex

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Имя теста | Accuracy | Precision | Recall | F-мера |
| Approximate F-test | 0.8620 | 0.7896 | 0.9870 | 0.8773 |
| Alexandern-Govern test | 0.8445 | 0.7673 | 0.9890 | 0.8641 |
| B-square test | 0.8450 | ***0.8120*** | 0.9890 | ***0.8918*** |
| Brown-Forsythe test | 0.8420 | 0.7659 | 0.9850 | 0.8618 |
| Box F-test | 0.8590 | 0.7836 | 0.9920 | 0.8756 |
| Cochran F-test | 0.8620 | 0.7887 | 0.9890 | 0.8776 |
| Fiducial Approach test | 0.8590 | 0.7836 | 0.9920 | 0.8756 |
| Generalized F-test | 0.8370 | 0.7584 | 0.9890 | 0.8585 |
| Johansen F-test | 0.8425 | 0.7645 | 0.9900 | 0.8627 |
| Modified Brown-Forsythe test | 0.8540 | 0.7779 | 0.9910 | 0.8716 |
| Modified Generalized F-test | 0.8460 | 0.7695 | 0.9880 | 0.8651 |
| Modified Welch Testt | 0.8290 | 0.7492 | 0.9890 | 0.8526 |
| One Stage test | 0.7245 | 0.6463 | 0.9910 | 0.7823 |
| One Stage Range test | 0.7300 | 0.6501 | ***0.9960*** | 0.7867 |
| Parametric Bootstrap test | 0.8510 | 0.7742 | 0.9910 | 0.8693 |
| Permutation F-test | 0.8805 | 0.8613 | 0.9070 | 0.8836 |
| Revised generalized F-test | ***0.8830*** | 0.8634 | 0.9100 | 0.8861 |
| Scott-Smith Test | 0.8710 | 0.7992 | 0.9910 | 0.8848 |
| Welch-Aspin test | 0.8485 | 0.7704 | 0.9930 | 0.8676 |
| Welch F-test | 0.8475 | 0.7713 | 0.9880 | 0.8663 |

На основании проведенных расчетов можно сказать следующее:

- Наибольшее количество правильно принятых гипотез о равенстве средних зарегистрировано у One Stage Range test [2] (996 из 1000) – соответственно, у данного теста наибольший показатель чувствительности;

- Наибольшее количество правильно отвергнутых гипотез о равенстве средних наблюдается у Revised generalized F-test (856 из 1000), и у данного теста наибольший показатель общей точности;

- Наибольшая специфичность (доля правильно принятых гипотез о равенстве средних из всего объема принятых гипотез о равенстве) и наибольшее качество у F-меры наблюдается у B-square test [3].

Таким образом, можно сделать вывод, что на малых выборках при малых значениях средней и дисперсии большую точность как и наименьшую вероятность допустить ошибку второго рода обеспечивает Revised generalized F-test, наименьшую вероятность допустить ошибку первого рода обеспечивает One Stage Range test, а наиболее точны выводы о правильности принятой гипотезы у B-square test.

Однако, данные результаты нуждаются в дальнейшем обобщении на случай многих групп с разными показателями средней и дисперсии.

ЛИТЕРАТУРА

1. Cavus, M., Yazıcı, B. Testing the equality of normal distributed and independent groups’ means under unequal variances by doex package
2. Chen, S.Y., Chen, H.J. A Range Test for the Equality of Means when Variances are Unequal / American Journal of Mathematical and Management Sciences. 2000. № 20. Р. 145-170.
3. Ozdemir, A. F., Kurt, S. One-way fixed effect analysis of variance under variance heterogeneityand a solution proposal / Selcuk Journal of Applied Mathematics. 2016. № 2(7). Р. 81–90