

**Об условиях применения критериев для проверки гипотез о нормальности  
распределения в прикладных исследованиях**

*Дьячкова Екатерина Николаевна*

*Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова,  
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия  
*E-mail: e.dyachkova.98@gmail.com*

В руководствах по прикладной статистике при использовании статистических критериев предлагается проводить предварительное тестирование данных на нормальность распределения [1], [2]. Мы рассмотрим этот вопрос, используя инструменты, которые были недоступны авторам пособий по прикладной статистике. Главным методом исследования будут оценки методом Монте-Карло, дополнительными — современные прикладные работы по количественной генетике и кардиологии. Доклад состоит из трёх частей. Первая часть доклада посвящена теоретическим аспектам проверки гипотезы о нормальности распределения. Мы покажем, что по результатам эмпирического исследования мощности критериев для проверки гипотез о нормальности распределения лидером оказался критерий Шапиро–Уилка. Далее мы покажем, что округление данных может влиять на результаты применения критериев нормальности. Во второй части доклада мы обратимся к прикладным данным. Мы обсудим результаты работы из области количественной генетики [3], в которой на данных Биобанка Великобритании показано, что рост взрослого человека, который исторически признавался нормальным [4], [5], [6], разумно считать логнормально распределённой величиной. Далее мы продемонстрируем результаты систематического обзора статей. В обзор вошли 356 статей за 2021 год из двух ведущих российских журналов по кардиологии: «Рациональная фармакотерапия в кардиологии» и «Российский кардиологический журнал». Оказалось, что в половине статей, содержащих статистический анализ, используется предварительная проверка данных на нормальность. При этом в 62 % статей, где была проведена проверка распределения данных на нормальность, эта проверка была проведена некорректно. В последней части доклада мы рассматриваем альтернативные подходы к оценке отклонения данных от нормальности. Вначале мы изучили асимптотическую сходимост рядов Эджворта. Оказалось, что на интересующих нас выборках до 100 элементов, которые нас интересуют, ошибка в оценке отклонения от нормальности слишком велика. Поэтому мы перешли к модели, где данные извлекаются из смеси нормального и логнормального распределения, а к полученным выборкам применяется критерий Стьюдента. В рамках этой модели мы сделали вывод, что при проверке одновыборочной гипотезы для контроля ошибки первого рода достаточно использовать показатели асимметрии.

**Источники и литература**

- 1) Берестнева О. Г., Марухина О. В., Шевелев Г. Е. Прикладная математическая статистика //Томск: Изд-во Томского политех. ун-та. – 2012.
- 2) Гланц С. Медико-биологическая статистика. Электронная книга. – 1999.
- 3) Slavskii, S. A., Kuznetsov, I. A., Shashkova, T. I., Bazykin, G. A., Axenovich, T. I., Kondrashov, F. A., & Aulchenko, Y. S. (2021)., The limits of normal approximation for adult height. European Journal of Human Genetics, 29(7), 1082-1091.

- 4) Snedecor GW., Statistical Methods: By George W. Snedecor and William G. Cochran. Iowa State University Press; 1989. 503 p.
- 5) Devore JL, Berk KN., Modern Mathematical Statistics with Applications. Springer, New York, NY; 2012.
- 6) Wright S., Evolution and the genetics of populations. Vol. 1. Genetic and biométrie foundations. London and Chicago: University of Chicago Press.; 1968.