И.В. Слезко
Тюменский государственный университет, г. Тюмень

Е.Б. Слезко
«Физико-математическая школа» Тюменской области, г. Тюмень

УДК 519.234.3

СТАТИСТИЧЕСКИЙ АНАЛИЗ В ПРОЕКТНОЙ ДЕЯТЕЛЬНОСТИ ШКОЛЬНИКА

Аннотация. В данной статье рассмотрен вопрос об использования методов статистического анализа в проектной деятельности школьника на примере статистической проверки оптимальности алгоритма Евклида при поиске наибольшего общего делителя.

Ключевые слова: алгоритм Евклида, статистический непараметрический критерий, многофункциональный статистический критерий.

В современных динамичных условиях развития и популяризации инженерных наук на первое место выходит проектная деятельность школьников. Она затрагивает различные области научного знания: гуманитарные и социальные, естественные науки, физику, математику. Проектная деятельность часто предполагает проведение эксперимента – натурного или вычислительного. Любой эксперимент подразумевает не только сбор, но и обработку полученной информации. Чтобы сделать достоверные выводы о значении той или иной величины, об изменении того или иного параметра применяются методы статистического анализа.

Заметим, что при использовании методов статистического анализа, нужно уметь интерпретировать полученные результаты, что иногда оказывается достаточно сложным. Практика подсказывает, что чем проще методы математической обработки, тем более надежными и осмысленными являются результаты. При выборе статистических критериев в данном исследовании будем руководствоваться их простотой и практичностью. Наиболее простыми являются непараметрические статистические критерии, которые основаны на оперировании частотами или рангами и не содержат в расчетной формуле параметров распределения (например, Т-критерий Вилкоксона). Однако такие критерии иногда оказываются маломощными и не позволяют достоверно выявить различия. В этом случае можно применить многофункциональные критерии, чтобы избежать обращения к параметрическим. Таким критерием является, например, достаточно простое и практичное угловое преобразование Фишера.

В качестве примера обработки результатов эксперимента, проведенного в ходе выполнения работы в рамках проектной деятельности школьника, рассмотрим статистическую проверку гипотезы об оптимальности алгоритма Евклида при поиске наибольшего общего делителя (НОД). Поиск НОД двух и более чисел требуется осуществлять при решении многих арифметических задач.

Наибольшим общим делителем нескольких чисел называется наибольшее число, на которое делятся все эти числа [1]. Существует два способа нахождения НОД. Самым распространенным является способ разложения на простые множители. Однако нахождение таких разложений может оказаться очень трудным занятием для больших чисел. Существует другой метод для нахождения наибольшего общего делителя, который не использует подобных разложений. Таким методом является так называемый алгоритм Евклида – метод последовательного деления [2]. Он реализуется достаточно просто, если исследователь умеет делить одно число на другое. Заметим, что алгоритм Евклида назван в честь ученого потому, что самое раннее известное его появление — в книге VII «Начал» [3].

Для оценки трудоемкости различных способов поиска НОД двух и более чисел применим метод хронометрирования, суть которого заключается в подсчете времени, затрачиваемого на решение одного примера одним способом.

Испытуемым было предложено несколько примеров для отыскания НОД двух и более чисел. Пример решения тестового примера и результаты хронометрирования (время – в секундах) приведены в таблице1.

Таблица 1.

<table>
<tr><td colspan="12">НОД(4199; 2431; 6936)=17</td></tr>
<tr><td colspan="6">Разложение на множители</td><td colspan="6">Алгоритм Евклида</td></tr>
<tr><td>4199</td><td>17</td><td>2431</td><td>11</td><td>6936</td><td>2</td><td colspan="3">НОД(4199; 2431)=221</td><td colspan="3">НОД(6936;221)</td></tr>
<tr><td>247</td><td>13</td><td>221</td><td>13</td><td>3468</td><td>2</td><td colspan="3">4199:2431=1 (ост. 1768)</td><td colspan="3">6936:221=31 (ост.85)</td></tr>
<tr><td>19</td><td>19</td><td>17</td><td>17</td><td>1734</td><td>2</td><td colspan="3">4199=2431·1+1768</td><td colspan="3">6936=221·31+85</td></tr>
<tr><td>1</td><td></td><td>1</td><td></td><td>876</td><td>3</td><td colspan="3">2431:1768=1 (ост.663)</td><td colspan="3">221:85=2 (ост.51)</td></tr>
<tr><td></td><td></td><td></td><td></td><td>289</td><td>17</td><td colspan="3">2431=1768·1+663</td><td colspan="3">221=85·2+51</td></tr>
<tr><td></td><td></td><td></td><td></td><td>17</td><td>17</td><td colspan="3">1768:663=2(ост. 442)</td><td colspan="3">85:51=1 (ост.34)</td></tr>
<tr><td></td><td></td><td></td><td></td><td>1</td><td></td><td colspan="3">1768=663·2+442</td><td colspan="3">85=51·1+34</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="3">663:442=1 (ост.221)</td><td colspan="3">51:34=1 (ост.17)</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="3">663=442·1+221</td><td colspan="3">51=34·1+17</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="3">442:221=2 (ост.0)</td><td colspan="3">34:17=2 (ост.0)</td></tr>
<tr><td colspan="6">НОД=17</td><td colspan="6">НОД= 17</td></tr>
<tr><td colspan="6">Время</td><td colspan="6">Время</td></tr>
<tr><td colspan="2">491</td><td colspan="2">425</td><td colspan="2">316</td><td colspan="2">230</td><td colspan="2">215</td><td colspan="2">415</td></tr>
<tr><td colspan="2">463</td><td colspan="2">425</td><td colspan="2">286</td><td colspan="2">485</td><td colspan="2">302</td><td colspan="2">192</td></tr>
<tr><td colspan="2">117</td><td colspan="2">459</td><td colspan="2">900</td><td colspan="2">493</td><td colspan="2">748</td><td colspan="2">200</td></tr>
<tr><td colspan="2">459</td><td colspan="2">150</td><td colspan="2">960</td><td colspan="2">277</td><td colspan="2">263</td><td colspan="2">860</td></tr>
<tr><td colspan="2">545</td><td colspan="2">336</td><td colspan="2">234</td><td colspan="2">430</td><td colspan="2">272</td><td colspan="2">191</td></tr>
<tr><td colspan="2">40</td><td colspan="2">390</td><td colspan="2">-</td><td colspan="2">300</td><td colspan="2">272</td><td colspan="2">-</td></tr>
</table>

По результатам таблицы сравним время, затрачиваемое на поиск НОД двумя методами. Можно сделать предположение, что при использовании алгоритма Евклида, время решения задачи уменьшается. Заметим, что различия в длительности попыток решения задачи в сторону уменьшения при использовании алгоритма Евклида зафиксированы при поиске НОД следующих чисел: (11687; 899), (323; 102), (1441; 1401), (69069; 13104), (4199; 2431; 6936). При поиске НОД чисел (94; 14) существенных различий не наблюдалось.

Для сопоставления показателей, измеренных в двух разных условиях на одной и той же выборке испытуемых, будем использовать такие статистические критерии, как Т-критерий Вилкоксона и угловое преобразование Фишера [4].

Т-критерий позволяет установить не только направленность изменений, но и их выраженность. С его помощью мы определяем, является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Критерий Фишера позволяет сопоставить две выборки по частоте встречаемости исследуемого эффекта. Заметим, что критерий Фишера применяем в том случае, если Т-критерий не позволил выявить различия.

Проверим достоверность различий при поиске НОД (4199, 2431, 6936) при заданном числе испытуемых n=17. Проверяем выполнение ограничений Т-критерия и, в соответствии с алгоритмом, рассчитываем эмпирическое значение критерия. Определяем сдвиги в замерах «до» и «после», устанавливаем направление типичного сдвига (в данном примере типичным направлением является отрицательное). В соответствии с этим формулируем гипотезы об интенсивности сдвига:

H_0: Интенсивность сдвига в сторону уменьшения времени решения задачи с помощью алгоритма Евклида не превосходит интенсивности сдвига в сторону его увеличения.

H_1: Интенсивность сдвига в сторону уменьшения времени решения задачи с помощью алгоритма Евклида превосходит интенсивность сдвига в сторону его увеличения.

Ранжируем сдвиги как единую выборку и находим сумму рангов в нетипичном направлении: $T_{эмп} = 35$. Критические значения критерия для разных уровней значимости p: $T_{кр}$=27, при p=0,01: $T_{кр}$=41, при p=0,05.

Сравниваем эмпирическое значение с критическим. Так как $T_{0.01} < \text{Т}_{\text{эмп}} < T_{0.05}$, то делаем вывод об отклонении основной гипотезы H_0. Но и альтернативную гипотезу H_1 принять не можем. Мощности критерия недостаточно, чтобы достоверно судить о различиях в уровне рассматриваемого признака.

Для дальнейшего анализа используем угловое преобразование Фишера. Под «эффектом» будем понимать уменьшение длительности времени решения задачи с помощью алгоритма Евклида. Сформулируем гипотезы:

H_0: доля лиц, которым требуется меньше времени на нахождение НОД с помощью алгоритма Евклида, чем разложением на простые множители, не больше, чем доля лиц, которым требуется меньше времени.

H_1: доля лиц, которым требуется меньше времени на нахождение НОД с помощью алгоритма Евклида, чем разложением на простые множители, больше, чем доля лиц, которым требуется меньше времени.

Определяем число испытуемых с «эффектом» и без «эффекта» и рассчитываем эмпирическое значение критерия:

	Число наблюдений	Процентная доля (%)	Угол
С «эффектом»	4	22,2	0,982
Без «эффекта»	14	77,8	2,160

$$\varphi^*_{\text{эмп}} = (2{,}16 - 0{,}982) \cdot \sqrt{\frac{4 \cdot 14}{4+14}} = 2{,}08.$$

Если $\varphi^*_{\text{эмп}} \geq \varphi^*_{\text{кр}}\ (2{,}08 > 1{,}64)$, то H_0 отклоняется. Достоверность различий принимается на уровне значимости *p*=0,019 (данное значение устанавливается по таблицам уровней значимости углового преобразования Фишера).

Таким образом, наблюдаемые различия в длительности попыток нахождения НОД различными методами не являются случайными, то есть различия между показателями во времени решения задач с использованием разных методов достоверны.

Аналогично установлена достоверность различий в остальных задачах.

В ходе выполнения проекта рассмотрен вопрос о нахождении НОД двух и более чисел. Изучены и реализованы такие методы нахождения НОД, как метод разложения на простые множители и метод последовательного деления - алгоритм Евклида.

В экспериментальной части работы проведено хронометрирование поиска НОД различными методами, результаты которого позволили предположить, что применение алгоритма Евклида является наиболее оптимальным способом поиска НОД. С помощью статистических критериев Т Вилкоксона и углового преобразования Фишера проверены гипотезы о достоверности различий между двумя методами поиска НОД. В результате проверки выявлены достоверные различия в длительности поиска НОД двумя методами.

Статически подтверждено, что метод последовательного деления наиболее эффективен, особенно в следующих случаях:

- меньшее число является делителем большего;
- числа, НОД которых находится, являются большими;
- числа, НОД которых находится, имеют в качестве делителей большие простые числа;
- при поиске НОД для более чем двух чисел (больших чисел).

СПИСОК ЛИТЕРАТУРЫ

1. Алфутова Н.Б., А.В. Устинов. Алгебра и теория чисел для математических школ. – М.: МЦНМО, 2002. – 264 с.
2. Виноградов И.М. Основы теории чисел. – Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика», 2003. – 176 с.
3. Начала Евклида. Книги VII-X / пер. с греч. – Москва-Ленинград: Госуд.изд-во технико-теоретической литературы, 1949. – 511 с.
4. Сидоренко Е.В. Методы математической обработки в психологии. - СПб: Социально-психологический центр, 1996. – 349 с.