

УДК 004.852

ГЕНЕРАЦИЯ СТАТИСТИЧЕСКИХ ДАННЫХ ДЛЯ ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ

Березин М. А., студент группы ИПБ-24, II курс
Научный руководитель: Березина Л.В., к. т. н.
Рыбинский государственный технический университет
имени П.А. Соловьева
г. Рыбинск.

Математическая статистика является одним из важнейших инструментов современного инженера. Она позволяет принимать обоснованные решения в условиях неопределенности, работать с массивами практических данных, выявлять закономерности и ошибки. Первичным навыком работы со статистическими данными является их предварительная обработка: оценка качества полученной выборки, отсев грубых ошибок (выбросов), группировка и структурирование, представление в удобном для дальнейшего анализа виде. При организации учебного процесса становится актуальным вопрос подбора числовых значений, обеспечивающих максимально эффективное освоение методик обработки практических данных. При этом, нужно учитывать несколько моментов.

Во-первых, случайные значения подлежащие обработке должны иллюстрировать реальную техническую задачу. В этом случае процесс обработки и полученные выводы становятся более осмысленными.

Во-вторых, не смотря на случайность выборки, её состав должен быть управляемым. Преподаватель должен иметь возможность задавать закон распределения статистических данных, ожидаемое среднее значение и дисперсию, регулировать наличие выбросов и т. д. Это сделает процесс освоения учебного материала поэтапным и управляемым.

В-третьих, для современных молодых людей, выросших в эпоху цифровых технологий, важна наглядность и вовлечённость в процесс. Процесс обучения, содержащий элементы игры, становится более позитивным и, следовательно, более эффективным.

Для решения перечисленных задач была разработана программа генерации наборов статистических данных.

Программа реализована на языке C++ с использованием графической библиотеки SFML-2.6.2. Имеется возможность поддержки программы в операционных системах: Windows 10 и Linux. Регистрационная форма под Linux создана с помощью Qt Creator, а под Windows — с помощью Windows Forms.

Работа программы начинается с фиксации данных пользователя (рис. 1): фамилии, имени, отчества, группы.

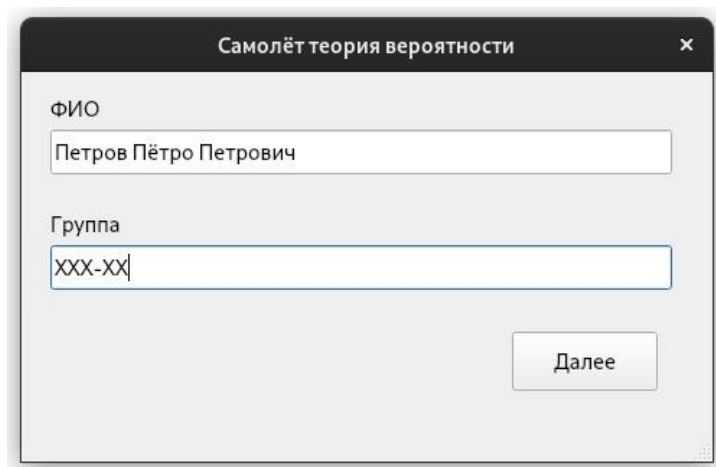


Рис. 1.

В программе моделируется процесс контроля удельного расхода топлива (кг/ч) двигателем самолета при полете на разных высотах, при различных скоростях в последовательные моменты времени. У студента есть возможность выбора определенных параметров высоты и скорости полета из нескольких вариантов. Выборочная совокупность значений удельного расхода топлива формируются на основе нормального закона распределения. Числовые характеристики распределения (математическое ожидание и среднее квадратическое отклонение удельного расхода) задаются в программе на основе реальных практических данных для существующих моделей летательных аппаратов. Процесс генерации значений сопровождается «игровым сценарием» (рис. 2): самолет взлетает, набирает высоту, совершает полет, приземляется. Предполагается, что данные расхода топлива фиксируются только в установившемся режиме полета.

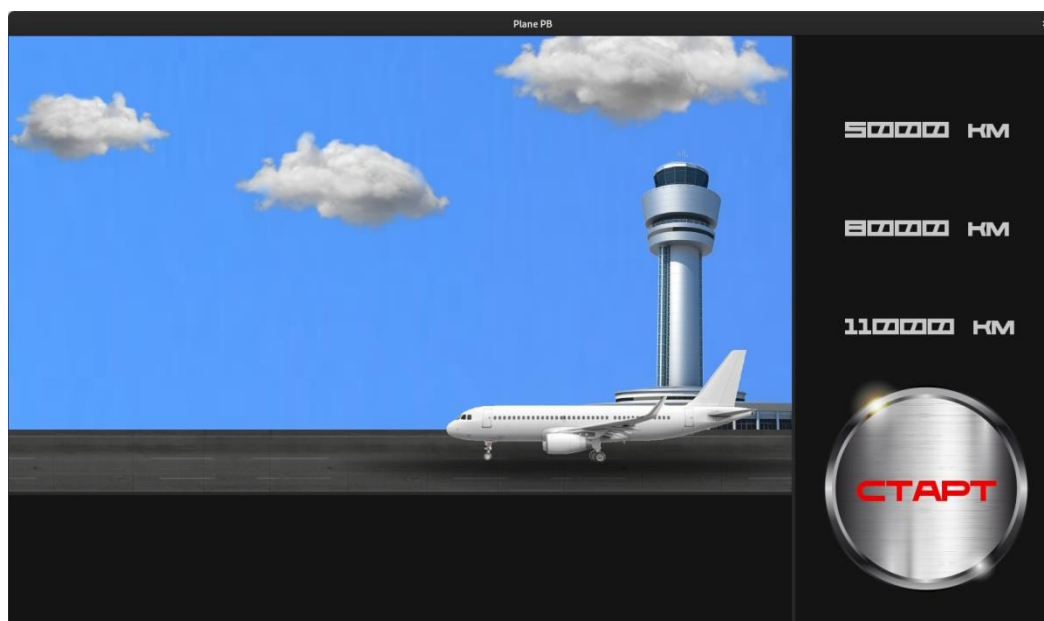


Рис. 2.

В полученную выборку вносятся несколько значений, имитирующих грубые ошибки.

Введение в выборочную совокупность «ошибочных» значений обусловлено следующими факторами:

1. Имитация реальных условий.

Практические данные никогда не являются идеальными. Всегда возможны неточности измерений, сбои приборов и влияние неучтенных факторов. Поэтому наличие выбросов приближает учебную задачу к реальной практике.

2. Демонстрация чувствительности выборочных оценок.

Присутствие в выборке грубых оценок позволяет наглядно показать искажение усредненных выборочных параметров.

3. Освоение методов обнаружения выбросов и корректировки выборочной совокупности.

Студенты получают возможность на практике применить методики выявления грубых ошибок, основанные на статистических критериях. А также научиться принимать обоснованные решения об удалении из совокупности числовых значений.

Величина значения, являющегося грубой ошибкой, в программе рассчитывается на основе правила «трех сигм». Согласно этому правилу, абсолютное отклонение нормально распределенной случайной величины от математического ожидания ($M(X)$) не превышает утроенного среднего квадратического отклонения ($S(X)$). Таким образом, после генерации выборки, наибольшее или наименьшее или оба значения заменяются на выбросы, полученные по формуле:

$$x_{max} = M(X) + 3,5 S(X) \text{ для наибольшего значения выборки,}$$

$$x_{min} = M(X) - 3,5 S(X) \text{ для наименьшего значения выборки.}$$

Таким образом, в ходе работы программы формируется текстовый файл с уникальным номером варианта, содержащий индивидуальные данные студента и набор из 100 значений, записанных в произвольном порядке. Для преподавателя сохраняется список студентов по группам с номерами вариантов.

Список литературы:

1. ГОСТ Р 8.736–2011 Государственная система обеспечения единства измерений. Измерения прямые многократные. Методы обработки результатов измерений. Основные положения. — М.: ФГУП Стандартинформ, 2013.
2. Ганичева, А. В. Прикладная статистика / А. В. Ганичева. — 4-е изд., стер. — Санкт-Петербург : Лань, 2023. — 172 с.