УДК 004.02 + 311.15

Кулакова Ирина Михайловна,
к.т.н., доцент, Ангарский государственный технический университет,
e-mail: iyelkina@mail.ru
Ерофеев Ермак Владимирович,
студент, Ангарский государственный технический университет,
e-mail: ermak.080 @mail.ru

ПРИЛОЖЕНИЕ ДЛЯ ВИЗУАЛИЗАЦИИ ДАННЫХ С ПОМОЩЬЮ BOXPLOT

Kulakova I.M., Erofeev E.V.

AN APPLICATION FOR VISUALIZING DATA USING BOXPLOT

Аннотация В данной работе анализируется график «Boxplot» как мощный инструмент для визуализации распределения данных, который позволяет оценить центральную тенденцию, разброс и выбросы. Рассматриваются различные параметры настройки графика, включая выбор границ ящика, настройки усов и отображение медианы и среднего арифметического. Понимание этих настроек способствует более точной интерпретации данных и принятию обоснованных решений в анализе.

Ключевые слова: Boxplot, визуализация данных, анализ данных.

Abstract. In this paper, we analyze the Boxplot graph as a powerful tool for visualizing the distribution of data, which allows us to estimate the central trend, spread, and outliers. Various graph settings are considered, including the selection of box borders, mustache settings, and the display of the median and arithmetic mean. Understanding these settings helps to interpret data more accurately and make informed decisions in the analysis.

Keywords: Boxplot, data visualization, data analysis.

График BOXPLOT является мощным инструментом для визуализации распределения данных. Он предоставляет информацию о центральной тенденции, разбросе и возможных выбросах в наборе данных [1]. Графики строятся с помощью разработанного авторами приложения, которое предоставляет гибкие инструменты для настройки графиков, что позволяет адаптировать их под конкретные задачи анализа. Это делает график не только визуально привлекательным, но и мощным средством для статистического анализа. Приложение было разработано в среде Embarcadero RAD Studio на языке Delphi 12.

Структура графика включает нескольких ключевых элементов: ящик – отображает интерквартильный диапазон (IQR), который включает в себя 1-й и 3-й квартили. Усы – представляют данные за пределами ящика и могут указывать на различные статистические параметры. Центральная линия – обозначает медиану. Центральная точка – представляет среднее арифметическое. Точки выбросов – значения, которые значительно отклоняются от других данных.

В разработанном приложении пользователь может выбрать, что будет обозначать границы ящика, с помощью различных опций:

- 1-3 квартиль. Этот выбор позволяет сосредоточиться на интерквантильном диапазоне, который показывает, где находится средняя часть данных. Это полезно для выявления разброса и центра данных (рис. 1 а, б, в, г).
- среднее арифметическое +/- стандартное отклонение. Этот вариант позволяет учитывать влияние крайних значений на распределение, что может

быть полезно в случаях, когда данные имеют нормальное распределение (рис. 1 д, е, ж).

Настройки усов могут значительно влиять на интерпретацию графика: Вариант от минимального до максимального значений предоставляет полное представление о диапазоне данных, включая крайние значения. Однако он может скрыть выбросы (рис. 1 б). Если за длину принимается доверительный интервал, то это позволяет пользователю задавать уровень доверия, что дает возможность учитывать неопределенность в оценках (рис. 1 а). Межквартильный размах относительно 1 и 3 квартилей учитывает разброс данных и позволяет пользователю регулировать настройки чувствительности к выбросам (рис. 1 в). Диапазон относительно среднего арифметического, кратный стандартному отклонению, позволяет учитывать вариативность данных и предоставляет более широкий взгляд на распределение (рис. 1 г).

Отображение медианы позволяет визуально оценить центр распределения данных (рис. 1 а, б, в, г, ж). Центральная точка позволяет сравнить среднее с медианой, что может быть полезно для выявления асимметрии в данных (рис. 1 д, е, ж).

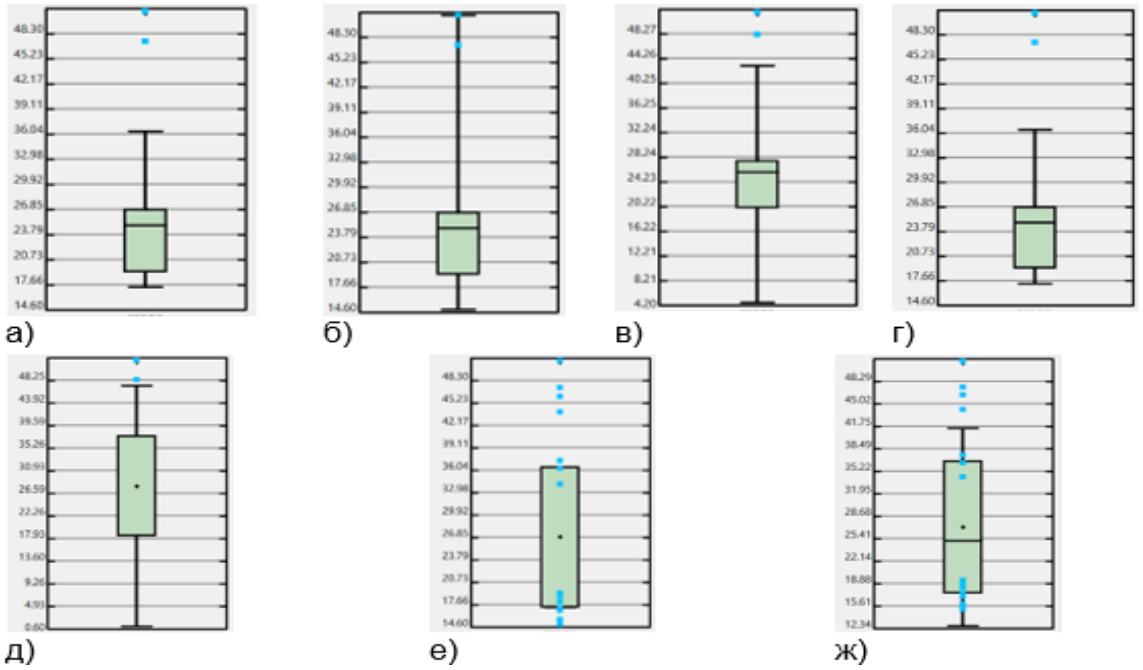

Рисунок 1 – Визуализация данных с помощью BOXPLOT

Таким образом, график BOXPLOT является универсальным инструментом для анализа данных, и его настройки могут значительно повлиять на интерпретацию результатов. Понимание того, как различные настройки влияют на визуализацию, поможет исследователям и аналитикам делать более обоснованные выводы и принимать решения на основе данных.

ЛИТЕРАТУРА

1. Как понимать Boxplot – Текст: электронный. – URL: https://www.tidydata.ru/boxplot (дата обращения: 11.02.2025).