

Особенности реализации SQL-запросов при регрессионном анализе данных многолетних наблюдений гидрометеостанции Биробиджан

Размахнина Анна Николаевна

*Приамурский государственный университет имени Шолом-Алейхема
студент*

Глаголев Владимир Александрович

*Приамурский государственный университет имени Шолом-Алейхема
к.г.н., доцент кафедры информационных систем, математики и методик
обучения*

Аннотация

В статье показано практическое использование SQL-запросов при выполнении регрессионного анализа многолетних наблюдений погодных условий и метеорологической пожарной опасности на примере данных гидрометеостанции Биробиджан с 1960 по 2009 гг., а также рассчитаны характеристики линейных трендов для каждого месяца пожароопасного сезона.

Ключевые слова: SQL, запрос, линейных тренд, пожарной опасность

Features of the implementation of SQL queries in regression analysis of long-term data of a weather station Birobidzhan

Razmakhnina Anna Nikolaevna

*Sholom-Aleichem Priamursky State University
Student*

Glagolev Vladimir Aleksandrovich

*Sholom-Aleichem Priamursky State University
Associate Professor of the Department of Information Systems, Mathematics and
teaching methods*

Abstract

The article shows the practical use of SQL queries when performing a regression analysis of long-term observations of weather conditions and meteorological fire danger in the example data of a weather station Birobidzhan from 1960 to 2009 and calculated characteristics of the linear trends for each month of the fire season.

Key words: SQL, query, linear trend, fire hazard

В настоящее время большое внимание уделяется анализу и прогнозу показателей метеорологической пожарной опасности на основе многолетних наблюдений за погодными условиями (50 и более лет). Для оперативного

построения прогноза необходимо выполнять множество запросов к большим массивам данных и обрабатывать полученные результаты вручную, либо в сочетании с современными вычислительными и программными средствами анализа. В работе рассмотрен способ выполнения статистического анализа операторами структурированного языка запросов *structured query language* (SQL) на сервере баз данных MySQL 5.11.

Изначально, язык запросов SQL состоял из небольшого количества операторов, позволяющих создавать таблицы, добавлять в них новые записи и извлекать существующие, удалять таблицы и изменять их структуру, в соответствии с заданными условиями. Со временем структура языка запросов SQL была переориентирована на визуализацию конечного результата обработки данных, а не на процесс их получения, хотя язык запросов SQL является мощным инструментом работы с базой данных [1].

Целью работы является разработка и практическая реализация SQL-запросов при выполнении регрессионного анализа дневной температуры воздуха и лесопожарного показателя пожарной опасности В.Г. Нестерова [2] на примере данных гидрометеостанции Биробиджан с 1960 по 2009 гг.

Корреляционно-регрессионный анализ проводится по следующему алгоритму [3]: учитываются цели и поставленные задачи исследования зависимости, задается зависимая и независимая переменная; определяются значения признаков, формируется выборочная совокупность значений; методом наименьших квадратов находятся оценки параметров уравнения регрессии; определяется теснота связи между изучаемыми признаками; оценивается значимость уравнения связи, его параметров и показателей тесноты связи; уравнение линейного тренда (регрессии) имеет следующий вид: $y = a + b \cdot x$.

Следуя алгоритму, производим вычисление характеристик линейной регрессии и коэффициентов корреляции для каждого месяца пожароопасного сезона [4]. Набор данных включает дневную температуру воздуха и лесопожарный показатель В.Г. Нестерова. Лесопожарный показатель пожарной опасности рассчитывается как произведение дневной температуры воздуха на разность дневной температуры воздуха и точки росы, измеренные и рассчитанные в 13-15 часов дня местного времени.

Была спроектирована и создана реляционная база данных Meteodata, содержащая 7 таблиц (April, May, June, July, August, September, October). Каждая таблица имеет одинаковую структуру: id—первичный ключ; temp, tros— дневная температура воздуха и точки росы, тип данных float; lpz— лесопожарный показатель пожарной опасности, тип данных float.

Для вычисления линейной регрессии использован метод наименьших квадратов в результате были созданы следующие SQL-запросы, представленные ниже на рис 1.

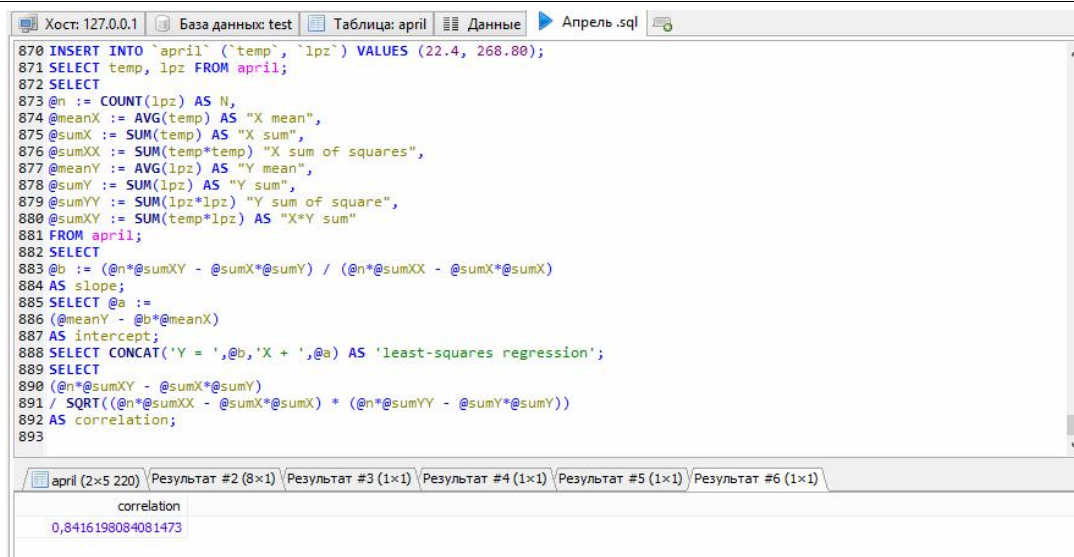


Рисунок 1. SQL-запрос для таблицы «April»

В SQL-запросе обозначим, дневную температуру воздуха – переменная @X, а показатель пожарной опасности – переменная @Y. Произведем вычисление величин, необходимых для расчета коэффициента корреляции: количество @n, средние значения @meanX, @meanY, суммы @sumX, @sumY и суммы квадратов @sumXX, @sumYY, для обеих переменных, а также сумму произведений для каждой переменной @sumXY. Найдем угловые коэффициенты корреляции @b, @a и уравнение корреляции ('Y = @b,'X + @a). Используя полученные величины, находим коэффициент корреляции:

$$\frac{(@n * @sumXY - @sumX * @sumY)}{\sqrt{((@n * @sumXX - @sumX * @sumX) * (@n * @sumYY - @sumY * @sumY))}}$$

В результате работы были вычислены коэффициенты корреляции для каждого месяца пожароопасного сезона.

Таблица 1

Месяц	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь
Количество переменных	870	1235	1143	1137	1079	1142	955
a	19,57	22,81	20,77	22,25	15,73	11,81	17,05
b	-64,86	-146,22	-237,03	-358	-183,95	-39,35	-43,05
Детерминации R2	0,7056	0,6724	0,5329	0,5329	0,3364	0,3025	0,6084
Коэффициент Корреляции R	0,84	0,82	0,73	0,73	0,58	0,55	0,78

После проведения вычисления линейной регрессии метеорологических данных для города Биробиджана, получили высокие коэффициенты

корреляции. Что говорит о сильной связи между использованными линейными параметрами.

Сравнивая коэффициенты корреляции, приходим к выводу, что самый высокий коэффициент корреляции выявлен у апреля, наименьший у сентября. Получаем, что пожарная опасность в различные периоды отличается, но при этом имеет высокие показатели.

Библиографический список

1. SQL. URL: <https://ru.wikipedia.org/wiki/SQL> (дата обращения: 23.11.2016).
2. Нестеров В.Г., Гриценко М.В., Шабунина Т.А. Использование температуры точки росы при расчете показателя горимости леса // Метеорология и гидрология. 1968. № 9. С. 102–104
3. Корреляционно-регрессионный анализ. URL: <http://allrefs.net/c23/45sg5/p14/> (дата обращения: 23.11.2016).
4. Вычисление линейной регрессии и коэффициентов корреляции // <http://ooportal.ru/> URL: <http://ooportal.ru/> (дата обращения: 23.11.2016).