состояния кинематических узлов и механизмов, на пути практического построения которых имеется ряд ещё нерешенных теоретических и практических задач, рассмотрению которых будет посвящено дальнейшее исследование.

1. Автомобильный парк России [Электронный ресурс] // Аналитическое агентство «Автостат». – Режим доступа: <https://www.kommersant.ru/doc/4267978> – 25.02.2020.
2. Расчет ёмкости рынка автокомпонентов и запчастей для легковых автомобилей России. [Электронный ресурс] //INFO-PARTS.RU Режим доступа: <https://info-parts.ru/all-publications.html?start=40> – 24.09.2020.
3. Генкин, М.Д. Виброакустическая диагностика машин и механизмов /М.Д. Генкин, А.Г. Соколова – М.:Машиностроение, 1987. – 288 с.
4. Балицкий, Ф.Я. Виброакустическая диагностика зарождающихся дефектов /Ф.Я. Балицкий, М.А. Иванова, А.Г. Соколова, Е.И. Хомяков. – М.: «Наука», 1984. – 120 с
5. Вибродиагностика. Неразрушающий контроль. Справочник /Ф.Я. Балицкий, А.В. Барков, Н.А. Баркова – М.: Машиностроение, 2005. - 829 с.
6. Суворов В. Н. Многоканальный виброизмерительный комплект К5101 //Приборы и системы. Управление, контроль, диагностика. 2006, №1. С. 36-37
7. Рыжов С.Н. Средства вибродиагностики технологического оборудования // Датчики и Системы. 2006, № 10. – С. 43-46
8. Дрейзин, В.Э. Акустическая диагностика автомобильных двигателей. Анализ возможностей и теоретические и практические предпосылки [Текст] / В.Э. Дрейзин, М.М. Касем, Д.С. Сабельников // Приборы и системы. Управление, контроль, диагностика, № 4, 2009. С.48-56.
9. Гетманов В. Г., Гридин В. Н., Царева Е. Б. Применение технологии обобщенного Фурье-анализа для цифровой обработки структурно-сложных нестационарных колебательных сигналов // Информационные технологии. 2005, № 9. С. 55-60.
10. Меркушева А. В. Вреямасштабные преобразования для анализа динамики спектра нестационарного сигнала в информационно-измерительных системах // Информационные технологии. 2005, № 2. С. 44-48
11. Дмитриев Е. В. Расчет естественного спектра и тренда процессов, сигналов, функций // Приборы и системы. Управление, контроль, диагностика. 2007, №10. С. 38-44
12. Аврамчук В.С. Метод определения неисправного цилиндра двигателя внутреннего сгорания на основе частотно-временного корреляционного анализа /В.С. Аврамчук, В.А. Фаерман // Приборы и системы. Управление, контроль, диагностика. 2018, №3. С. 37-43

Окатов Д.А., Минеева Т.А.

Парный корреляционный анализ статистической обработки данных в медицине

*Северный (Арктический) федеральный университет имени М. В. Ломоносова
(Россия, Архангельск)*

doi: 10.18411/lj-04-2021-19

Аннотация

Данная статья посвящена изучению парного корреляционного анализа с целью использования его в статистической обработке данных в медицине. В статье рассматривается понятие парного корреляционного анализа, его применение в выявлении связи летальности от дня распространения эпидемии коронавируса.

Ключевые слова: корреляция, медицина, анализ, статистическая обработка данных.

Abstract

This article is devoted to the study of pair correlation analysis in order to use it in statistical data processing in medicine. The article discusses the concept of pair correlation analysis, its application in identifying the relationship of mortality from the day of the spread of the coronavirus epidemic.

Keywords: correlation, the medicine, analysis, statistical data processing.

В настоящее время корреляционный анализ имеет широкое применение в выявлении оценки тесноты связи между исследуемыми величинами с целью получения выводов о зависимости между ними.

Парный корреляционный анализ или корреляция Пирсона – это метод, характеризующий тесноту линейной корреляционной связи между зависимой случайной величиной и независимой случайной величиной.

Рисунок 1 – Графическое представление корреляционной связи

Данный метод показывает влияние одного явления на другое и их связь в динамике. Подобного рода зависимости существуют, например, между величинами, находящимися в причинно-следственных связях друг с другом. Если два явления статистически достоверно коррелируют друг с другом и при этом есть уверенность в том, что одно из них выступает в качестве причины другого явления, то можно сделать вывод о наличии между ними причинно-следственной зависимости [2].

Впервые в научный оборот термин корреляция ввёл французский палеонтолог Жорж Кювье в XVIII веке.

Он разработал «закон корреляции» частей и органов живых существ, с помощью которого можно восстановить облик ископаемого животного, имея в распоряжении лишь часть его останков.

Термин «линейный коэффициент корреляции» или «коэффициент корреляции Пирсона» был введен Карлом Пирсоном [3].

Рисунок 2 – Жорж Кювье (слева) и Карл Пирсон (справа)

Данный метод обработки статистических данных весьма популярен в экономике, астрофизике и социальных науках, а также в психологии и социологии, хотя сфера применения корреляционного анализа обширна: контроль качества промышленной продукции, металловедение, агрохимия, гидробиология, биометрия и прочие. В различных прикладных отраслях приняты разные границы интервалов для оценки тесноты и значимости связи.

Рассмотрим конкретный пример зависимости летальности от распространения эпидемии коронавируса.

Существует достаточно много разновидностей формулы расчета коэффициента парной корреляции. В данном случае уместно использовать следующую формулу:

$$r_{xy} = \frac{\sum(x_i - x_{cp})(y_i - y_{cp})}{\sqrt{\sum(x_i - x_{cp})^2 \sum(y_i - y_{cp})^2}} \quad (1)$$

Ошибка коэффициента корреляции находится по формуле:

$$m = \begin{cases} \frac{1-r_{xy}^2}{\sqrt{n}}, n > 50 \\ \frac{1-r_{xy}^2}{\sqrt{n-1}}, n \leq 50 \end{cases} \quad (2)$$

С достаточно большой надежностью можно утверждать, что зависимость неслучайна, если численное значение r_{xy} превышает свою среднюю ошибку не менее чем в 3 раза.

$$\frac{r_{xy}}{m} \geq 3 \quad (3)$$

Для реализации нахождения коэффициента корреляции был разработан код на языке C#. В листинге 1 представлен код реализации алгоритма.

Листинг 1

Код программы на языке C#

```
using System;
using System.Collections.Generic;

namespace CorrelationAnalysis
{
    class Program
    {
        static void Main(string[] args)
        {
            List<double> X = new List<double>() { };
            List<double> Y = new List<double>() { };
            List<double> x_av = new List<double>() { };
            List<double> y_av = new List<double>() { };
            List<double> x_av2 = new List<double>() { };
            List<double> y_av2 = new List<double>() { };
            List<double> xy = new List<double>() { };
            List<double> xy2 = new List<double>() { };
            double sumX = 0;
            double sumY = 0;
            Console.WriteLine("Введите количество записей");
            int countrows = Convert.ToInt32(Console.ReadLine());
            Console.WriteLine("Введите данные первого столбца");
            for (int i = 0; i < countrows; i++)
            {
                X.Add(Convert.ToDouble(Console.ReadLine()));
                Console.WriteLine("Введите данные второго столбца");
                for (int i = 0; i < countrows; i++)
                {
                    Y.Add(Convert.ToDouble(Console.ReadLine()));
                    for (int i = 0; i < X.Count; i++)
                    {
                        sumX += X[i];
                        sumY += Y[i];
                    }
                }
            }
        }
    }
}
```

```

        double xav = sumX / countrows;
        double yav = sumY / countrows;
        for (int i = 0; i < countrows; i++) x_av.Add(X[i] - xav);
        for (int i = 0; i < countrows; i++) y_av.Add(Y[i] - yav);
        for (int i = 0; i < countrows; i++)
x_av2.Add(Math.Pow(x_av[i], 2.0));
        for (int i = 0; i < countrows; i++)
y_av2.Add(Math.Pow(y_av[i], 2.0));
        for (int i = 0; i < countrows; i++) xy.Add(x_av[i] * y_av[i]);
        for (int i = 0; i < countrows; i++) xy2.Add(x_av2[i] *
y_av2[i]);
        double sum_x2 = 0;
        double sum_y2 = 0;
        double sumxy = 0;
        for (int i = 0; i < countrows; i++)
        {
            sum_x2 += x_av2[i];
            sum_y2 += y_av2[i];
            sumxy += xy[i];
        }
        double Correlation = sumxy / Math.Sqrt(sum_x2 * sum_y2);
        for (int i = 0; i < countrows; i++)
        {
            if (i == 0) Console.WriteLine("X\tY");
            Console.WriteLine(Math.Round(X[i]) + "\t" +
Math.Round(Y[i], 2));
        }
        Console.WriteLine("Коэффициент корреляции = " +
Math.Round(Correlation, 4));
        double m = 0;
        if (countrows > 50) m = (1 - Correlation * Correlation) /
Math.Sqrt(countrows);
        else m = (1 - Correlation * Correlation) / Math.Sqrt(countrows
- 1);
        Console.WriteLine("Средняя ошибка коэффициента корреляции = "
+ m);
        if ((Correlation / m) >= 3) Console.WriteLine("Связь
статистически значимая");
        else Console.WriteLine("Существование связи между изучаемыми
явлениями нельзя признать доказанным");
        double det = Correlation * Correlation;
        Console.WriteLine("Коэффициент детерминации = " + det);
    }
}
}

```

Для примера были взяты данные с сайта по статистике распространения эпидемии коронавируса в весенний период 2020 года в России [1].

Таблица 9

Данные зависимости летальности от распространения эпидемии коронавируса.

День распространения эпидемии коронавируса (x)	Летальность (y)
1	1,0
10	3,0
20	4,0
30	8,0
40	17,0
50	24,0
60	30,0
70	46,0

Результат работы программы, приведенной в листинге 1, представлен на рисунке 2.

X	Y
1	1
10	3
20	4
30	8
40	17
50	24
60	30
70	46

Коэффициент корреляции = 0,9566
 Средняя ошибка коэффициента корреляции = 0,03206008045929353
 Связь статистически значимая
 Коэффициент детерминации = 0,9151770000919879

Рисунок 3 – Результат обработки данных

Таким образом, коэффициент корреляции равен 0,9566. Связь положительная, сильная. Следовательно, между сроками распространения эпидемии коронавируса и летальностью имеется очень тесная зависимость.

Поскольку средняя ошибка коэффициента корреляции $\approx 0,032$, коэффициент корреляции превышает ошибку в 3 и более раз, а значит можно сделать вывод, что связь статистически значимая.

Поскольку летальность зависит не только от распространения эпидемии коронавируса, но и от множества других факторов, например болезни (туберкулез, гепатит А, ВИЧ, СПИД и др.), имеет место рассчитать коэффициент детерминации, показывающий, какая часть результативного признака вызвана факторным признаком. Формула для расчета коэффициента детерминации:

$$D = r_{xy}^2 \quad (4)$$

Коэффициент детерминации 0,915, близок к единице. Из этого следует, что в настоящее время влияние распространения эпидемии коронавируса велико, по сравнению с другими факторами, влияющими на летальность [5].

На основании исследования изучены понятие и принцип парного корреляционного анализа. На примере показана прямая и сильная зависимость между летальностью и распространением эпидемии коронавируса.

1. Coronavirus-Monitor [Электронный ресурс] / Коронавирус в России – Режим доступа: <https://coronavirus-monitor.info/country/russia/>, свободный (Дата обращения: 27.02.2021)
2. Студенту психологу [Электронный ресурс] / О корреляции простыми словами – Режим доступа: http://студентупсихологу.рф/stati/article_post/o-korrelyatsii-prostymi-slovami, свободный (Дата обращения: 28.02.2021)
3. FB [Электронный ресурс] / Жорж Кювье – Режим доступа: <https://yandex.ru/turbo/fb.ru/s/article/251309/uchenyiy-jorj-kyuve-biografiya-dostijeniya-otkryitiya-i-interesnyie-faktyi>, свободный (Дата обращения: 01.03.2021)
4. Metanit [Электронный ресурс] / Полное руководство по языку программирования C# 9.0 и платформе .NET 5 – Режим доступа: <https://metanit.com/sharp/tutorial/>, свободный (Дата обращения: 19.02.2021)
5. MathSemestr [Электронный ресурс] / Расчет коэффициента детерминации – Режим доступа: <https://math.semestr.ru/corel/prim2.php>, свободный (Дата обращения: 19.02.2021)