

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРИМЕНЕНИЯ МЕТОДОВ ПОСТРОЕНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Иордан В.И.,

научный руководитель канд. физ.-мат.наук Мансурова Т.П.

Сибирский Федеральный университет

Определение причинных связей имеет большое значение в математическом моделировании для научного предвидения, воздействия на процессы и изменения их. При построении математических моделей важно определить зависимость и независимость переменных описывающих тот или иной процесс. Форму связи между переменными можно определить линией регрессии. В исследованиях широко используются линейные регрессионные модели.

Цель работы: Освоить методы построения регрессионных моделей. Сделать сравнительный анализ вычисления коэффициентов в линейной регрессионной модели:

$$y = a_0 + a_1x.$$

Для получения регрессионной прямой существует два статистических метода [1, 2]:

1. Метод приведенной главной оси:
2. Метод наименьших квадратов:

Таблица №1

Уравнение линейной регрессии: $y = a_0 + a_1x$	
Метод приведенной главной оси (Метод 1)	Метод наименьших квадратов (Метод 2)
$a_0 = \bar{y} - r_{xy} \cdot \frac{S_y}{S_x} \cdot \bar{x}$ $a_1 = r_{xy} \cdot \frac{S_y}{S_x}$	$a_0, a_1 \text{ находят из системы:}$ $\begin{cases} n \cdot a_0 + \sum_{i=1}^n x_i \cdot a_1 = \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i \cdot a_0 + \sum_{i=1}^n x_i^2 \cdot a_1 = \sum_{i=1}^n x_i \cdot y_i \end{cases}$

Для реализации цели нашей работы, воспользуемся известными исследованиями:

1. Исследование концентрации СО на территории г. Красноярска [3];
2. Исследование выбросов промышленных предприятий в г. Красноярске [3];
3. Исследования катионного состава проб воды [2];
4. Исследование зависимости уровня знаний студентов от количества пропусков аудиторных занятий [4].

Выбранные нами исследования отличаются:

1. Объемом выборки;
2. Уровнем линейной связи.

В табл. №2 представлены результаты вычислений:

Таблица № 2

№ опыта	n (объем)	r_{xy}	a_1		a_0		Относительная погрешность
			Метод 1	Метод 2	Метод 1	Метод 2	
1	10	0,8	0,68	0,68	616,91	614,11	0,45%
2	16	0,37	0,08	0,08	21,85	21,91	0,27%
3	20	0,23	0,51	0,51	23,91	23,83	0,34%
4	26	-0,76	-0,08	-0,08	4,85	4,88	0,62%

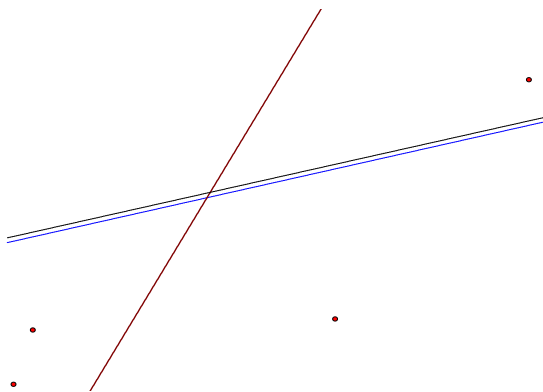


Рис.1. $n=16$, $r_{xy}=0,23$.

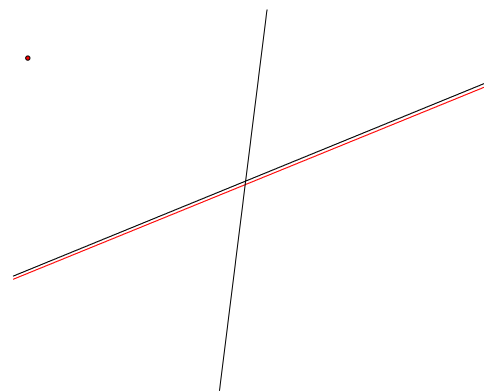


Рис.2. $n=20$, $r_{xy}=0,37$.

Анализируя табл. №2 и рис. 1, 2 можно сделать следующие выводы:

1. Оба метода построения линейных регрессионных моделей одинаково эффективны (a_0 совпадают, a_1 имеют недостоверные различия, менее 1%);
2. Результаты вычисления не зависят от объема выборки и уровня связи (жесткая - слабая, прямая - обратная).

Основные результаты:

1. Освоил методы построения регрессионных моделей: метод приведенной главной оси и метод наименьших квадратов;
2. Провел сравнительный анализ вычисления коэффициентов линейной регрессии.

Библиографический список:

1. Самнер Г. Математика для географов: Перевод с английского. – М: Прогресс, 1981. – 296 с.
2. Дэвис Дж.С. Статистический анализ данных в геологии (в 2-х книгах); М: «Недра».1990. -745с.
3. Р.Г. Хлебопрос, О.В. Тасейко, Ю.Д. Иванова, С.В. Михайлюта. Красноярск. Экологические очерки: монография. – Красноярск: Сибирский федеральный университет, 2012. – 136 с.
4. М.А. Тряпичкин, Т.П. Мансурова, Е.С. Закиева, Л.В. Климович. Влияние стимулирующих баллов за посещаемость на успеваемость студентов. Материалы IV научно-практической Интернет-конференции с международным участием «Информационные технологии и математическое моделирование в экономике, технике, экологии, образовании и торговле». – Красноярск – Гвадалахара, 2011. – С. 169-171.