МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОГРАФИИ

Карпиченко Александр Александрович

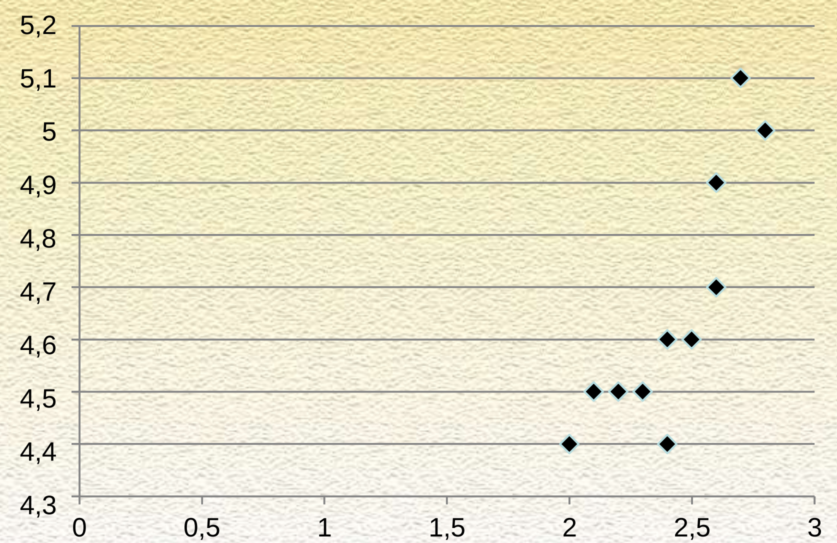
доцент кафедры почвоведения и земельных информационных систем

Литература

- elib.bsu.by
- Математические методы в географии: учебно-методическое пособие / Н. К. Чертко, А. А. Карпиченко. – Минск: БГУ, 2009.

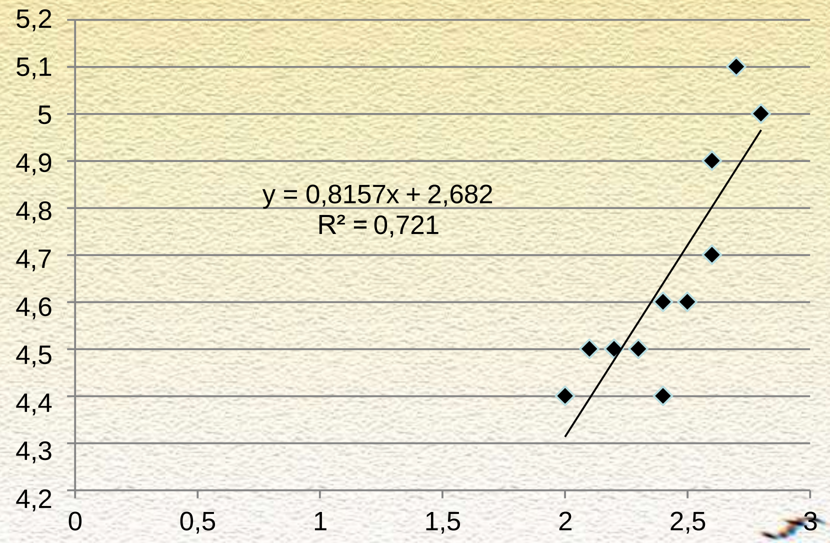
6. РЕГРЕССИОННЫЙ АНАЛИЗ

Логическим продолжением корреляционного анализа является регрессионный анализ, который развивает и углубляет представление о корреляционной связи. Если корреляционный анализ позволяет установить лишь форму и тесноту связи между случайными переменными, то регрессионный анализ математически описывает выявленную связь, т. е. дает возможность численно оценить одни параметры через другие.



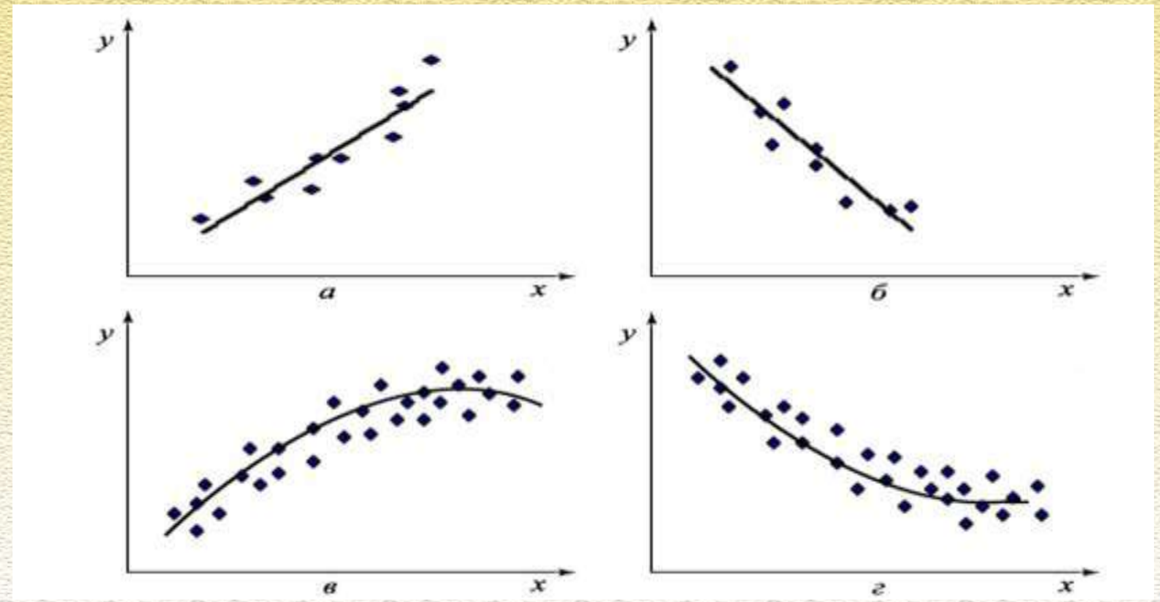
6. РЕГРЕССИОННЫЙ АНАЛИЗ

Составив и решив уравнения регрессии, можно произвести выравнивание эмпирических линий регрессии, т. е. моделировать наблюдаемую зависимость путем подбора функции, график которой представляет собой теоретическую линию регрессии. Если подобранная функция отражает сущность процесса или явления, то возможно прогнозирование значений признака за пределами сделанных наблюдений.



6. РЕГРЕССИОННЫЙ АНАЛИЗ

Подобно корреляции, регрессия может быть *парной* (простой) и *множественной*, по форме связи – *линейной* и *нелинейной*, по зависимости – *односторонней* (изменяется лишь один признак под влиянием другого) и *двусторонней* (изменяются оба признака под воздействием друг друга).



6. РЕГРЕССИОННЫЙ АНАЛИЗ

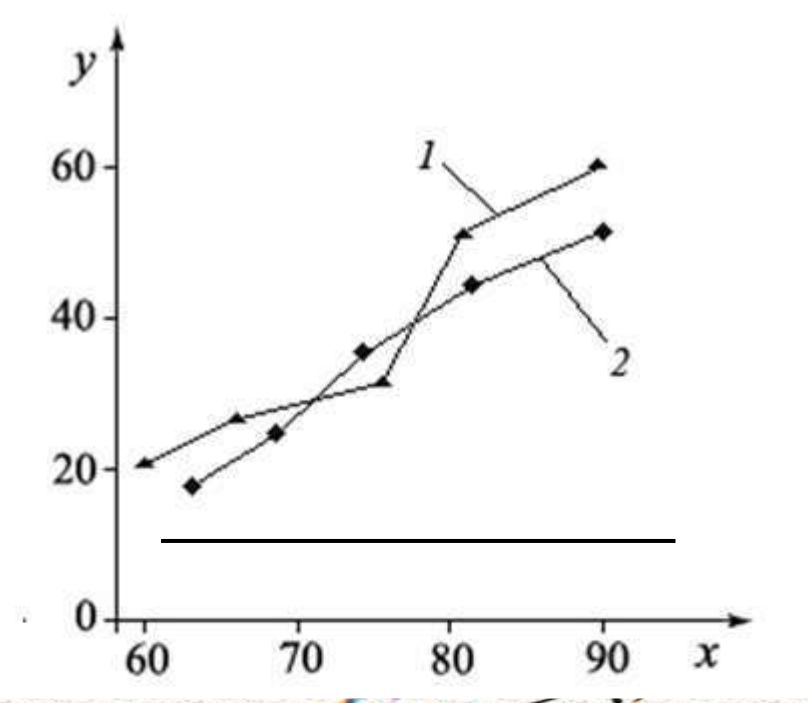
Регрессия выражается несколькими способами: построением эмпирических линий, составлением уравнения и затем – построением теоретических линий регрессии, а также с помощью коэффициента регрессии. Уравнение наиболее точно выражает зависимость между двумя переменными (*x*,*y*), если корреляция между ними близка к единице.

Регрессионный анализ возможен при наличии всего лишь нескольких пар сопряженных наблюдений, но при условии сильных связей между признаками ($r \geq 0,7$).

6. РЕГРЕССИОННЫЙ АНАЛИЗ

Точки эмпирических линий регрессии (1, 2) определяются как взвешенные средние арифметические, для невзвешенных рядов – как средние малых групп выборки. Вычислив координаты точек, наносим их на график и соединяем прямой; в результате получаются эмпирические линии регрессии. По графическому изображению можно предварительно сделать заключение о характере связи.

При полном отсутствии связи эмпирические линии располагаются параллельно осям графика. При полной связи между *x*, *y* (*r*=1) линии регрессии на графике, построенные по точкам эмпирических линий регрессии, совместятся.



6. РЕГРЕССИОННЫЙ АНАЛИЗ

Существует два способа составления уравнений регрессии: а) способ координат точек, с использованием двух-трех точек, расположенных на эмпирической линии (желательно в начале, середине и конце ее),– для тех случаев, когда расчет не требует большой точности; б) способ наименьших квадратов, более точный, так как для составления уравнения регрессии привлекаются все сопряженные наблюдения.

6.1. Линейная зависимость

Линейная регрессия на графике изображается в виде прямой так, чтобы точки эмпирической линии располагались по обе стороны ее и по возможности ближе к ней.

Известно следующее уравнение линейной регрессии:

$$y = ax + b$$

где *y* – значение зависимой переменной (признак); *x* – значение независимой переменной (фактор, влияющий на признак); *a* – коэффициент регрессии, показывающий степень зависимости между переменными (может быть также выражен тангенсом угла наклона линии регрессии к оси абсцисс); *b* – ордината линии, показывающая смещение начала прямой относительно начала координат.

6.1. Линейная зависимость

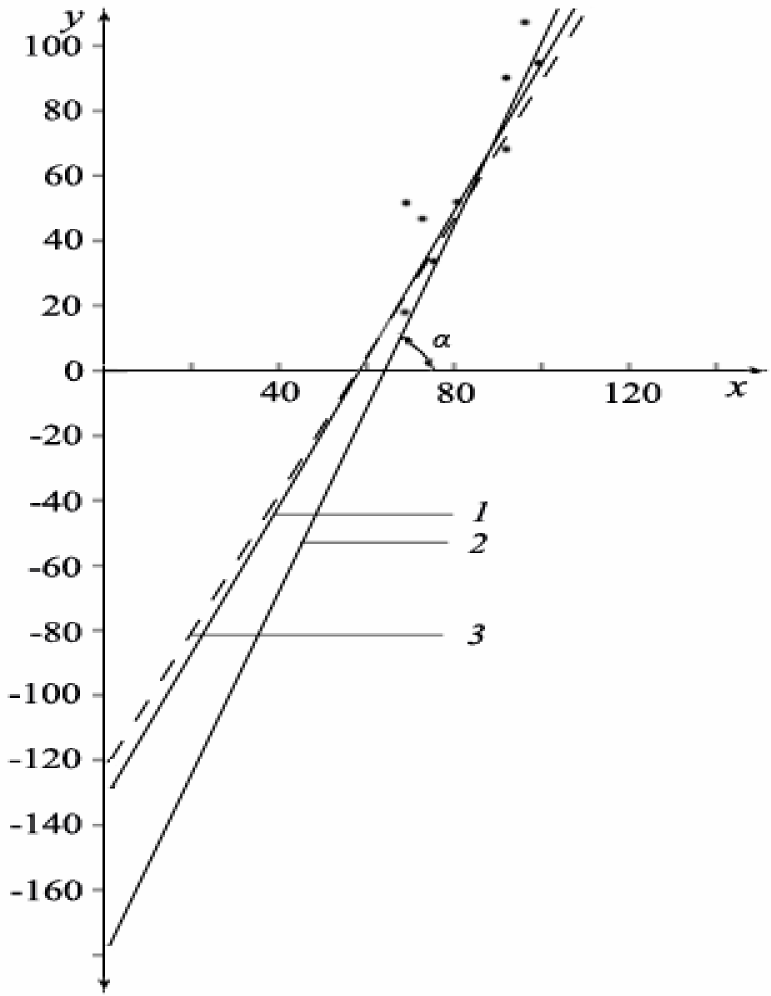


Рис. 6.2. Сравнение местоположения эмпирических линий (*1*, *2*) с теоретической (*3*) по зависимости содержания подвижного марганца y от гидролитической кислотности x ($\angle \alpha = 70°06' = \mathrm{tg}_x\ 2{,}76$): для эмпирических линий *1* $y = 2{,}30x - 130{,}9$; для *2* $y = 2{,}76x - 173{,}0$; $r_{0,99} = 0{,}87$.

6.1. Линейная зависимость

Степень совпадения теоретической и эмпирической линии регрессии можно проверить, используя критерий хи-квадрат. Если $\chi_ф^2 > \chi_т^2$, то можно указать на недостаточное соответствие теоретической линии регрессии эмпирическому ряду.

Составленные уравнения регрессии можно проверить на точность зависимости между переменными (x, y) не только по критерию хи-квадрат, но и по коэффициенту точности выравнивания линии r_1, отражающему степень приближения (соответствия) фактических данных наблюдения к вероятным. Этот коэффициент определяем по формуле, в которой $(y_ф - M_ф) = \alpha$ – отклонение индивидуальных вариант от общего среднего арифметического по y; $(y_ф - y_в) = \beta$ – отклонение индивидуальных экспериментальных вариант по y от расчетных по уравнению.

$$r_1 = \sqrt{\frac{\Sigma\alpha^2 - \Sigma\beta^2}{\Sigma\alpha^2}} = \sqrt{\frac{\Sigma(y_Ф - M_Ф)^2 - \Sigma(y_Ф - y_в)^2}{\Sigma(y_Ф - M_Ф)^2}}$$

6.1. Линейная зависимость

Расчет данных для определения точности выравнивания линии

y		α-отклонения		β-отклонения	
$y_ф$	$y_в$	$y_ф - M_ф$	$(y_ф - M_ф)^2$	$y_ф - y_в$	$(y_ф - y_в)^2$
18	27,8	–42	1764	–9,8	96,04
48	30,1	–12	144	17,9	320,41
42	34,7	–18	324	7,3	53,29
31	41,6	29	841	–10,6	112,36
56	60,0	–4	16	–4,0	16,00
84	76,1	24	576	7,9	62,41
56	76,1	–4	16	–20,1	404,01
68	78,4	4	16	–10,4	108,16
90	87,6	30	900	2,4	5,76
107	87,6	47	2209	19,4	376,36
$M_ф = 60$			Σ 6806		Σ 1554,80

6.1. Линейная зависимость

Принято считать: если $r_1 > 0{,}95$, то уравнение регрессии соответствует более точному положению линии на графике. При $r_1 < 0{,}95$ необходимо найти другую математическую зависимость. В приведенном примере $r_1 = 0{,}88 < 0{,}95$, поэтому следует подобрать другую математическую зависимость. Такие же выводы получены при проверке на точность зависимости между переменными по критерию хи-квадрат. Оба критерия оценки (χ^2, r_1) на точность выравнивания линии уравнения регрессии используются и для других форм регрессионной зависимости (гиперболической, параболической).

6.1. Линейная зависимость

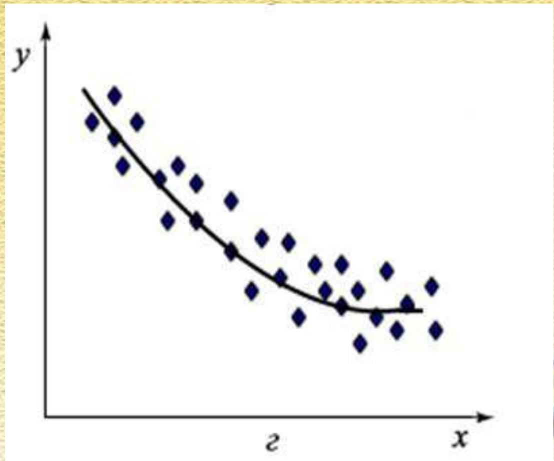
Ошибку уравнения регрессии можно определить по формуле, где *n* – число точек линии регрессии; *k* — число коэффициентов в уравнении регрессии (два плюс свободный член уравнения).

$$m = \sqrt{\frac{\Sigma(y - y')^2}{n-k}} = \sqrt{\frac{\sum(y_{ф} - y_{в})^2}{n-k}}$$

6.2. Гиперболическая зависимость

При нелинейной зависимости между аргументом и функцией, представляющая собой на графике кривую в виде гиперболы. Общее уравнение регрессии для гиперболической зависимости имеет вид:

$$y = a/x + b$$



где *x* – аргумент;

y – функция;

a и *b* – коэффициенты,

величину которых следует установить.

Правильный вид записи уравнения регрессии:

$$y = a/x + b;\ \eta_{0,95} = 0,84$$

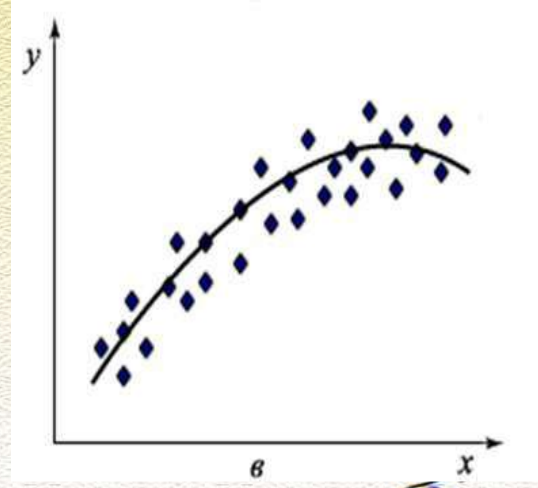
6.3. Параболическая зависимость

Общее уравнение параболы *n*-го порядка имеет вид

$$y = ax^n + bx^{n-1} + cx^{n-2} + \ldots + kx + l.$$

Если ограничиться второй ступенью независимой переменной величины *x*, будем иметь частный случай параболы второго порядка:

$$y = ax^2 + bx + c$$

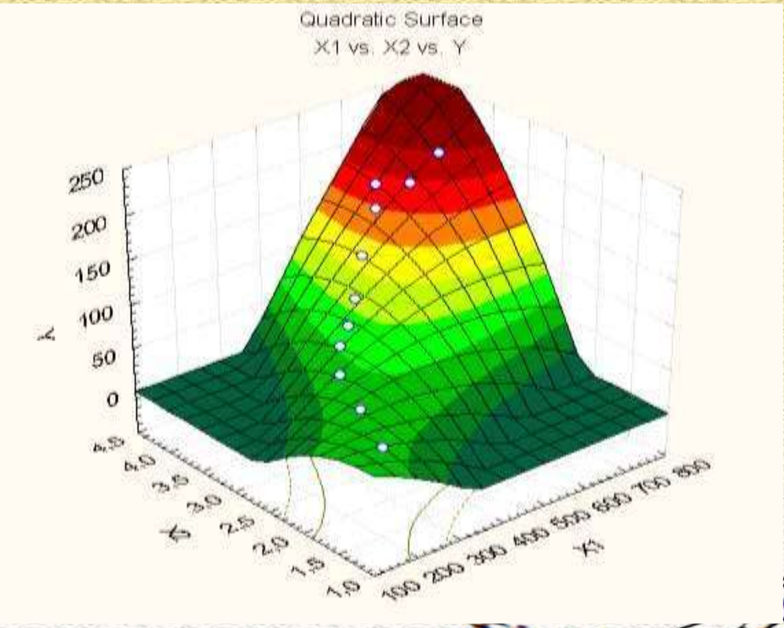


6.4. Множественная регрессия

Если при установлении зависимости между признаками используется больше одной независимой переменной, то применяют множественный регрессионный анализ. Проведение такого анализа возможно в следующих условиях: распределение зависимой переменной при различных значениях независимых должно быть близко к нормальному; дисперсия зависимой переменной при разных значениях признаков x должна считаться одинаковой.

6.4. Множественная регрессия

С увеличением числа признаков и в случаях нелинейной множественной регрессии необходимо использовать специализированное программное обеспечение. В простейшем случае, когда один признак зависит от двух факторов, общее уравнение линейной множественной регрессии имеет вид



$$y = a + bx + cz$$