«Оптика атмосферы и океана», **18**, № 1–2 (2005)

УДК 551.521.3

Н.Н. Щелканов

Обобщенный метод построения линейной регрессии и его применение для построения однопараметрических моделей аэрозольного ослабления

Институт оптики атмосферы СО РАН, г. Томск

Поступила в редакцию 12.11.2004 г.

Представлена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1X$ для общего случая, когда разброс точек в корреляционной связи величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами. Все известные выражения для коэффициентов регрессии оказались частными случаями полученной формулы. Предложены способы и приведены выражения для оценки среднеквадратических погрешностей измеряемых величин, входящих в формулу для вычисления коэффициентов регрессии, из экспериментальных данных.

Введение

При работе с экспериментальными данными часто возникает необходимость нахождения коэффициентов линейной регрессии между двумя физическими величинами. В большинстве случаев коэффициенты регрессии имеют конкретный физический смысл и для корректной интерпретации полученных результатов очень важно найти их значения наилучшим образом. Существует несколько формул для нахождения коэффициентов регрессии [1–3], но не для всех есть общее понимание, в каких случаях их следует использовать. В настоящее время отсутствует единый подход к нахождению коэффициентов линейной регрессии для общего случая, т.е. когда разброс точек в корреляционной связи между двумя величинами обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами.

Цель настоящей работы заключается в том, чтобы получить обобщенную формулу для вычисления коэффициентов линейной регрессии и применить ее для построения однопараметрических региональных моделей аэрозольного ослабления.

1. Постановка задачи

Рассмотрим две физические величины X_0 и Y_0, между которыми существует статистическая корреляционная связь. Предположим, что эта связь может быть описана линейной зависимостью

$$Y_0 = K_0 + K_1 X_0, \qquad (1)$$

а требуется найти коэффициенты регрессии K_0 и K_1, которые наилучшим образом отражают физическую взаимосвязь между ними.

Так как X_0 и Y_0 измеряются со случайными погрешностями, то на практике мы имеем дело с величинами X и Y, для которых уравнение регрессии запишется в виде

$$Y = K_0 + K_1 X. \qquad (2)$$

Запись уравнений (1) и (2) с одинаковыми коэффициентами регрессии говорит о том, что последние не должны зависеть от случайных погрешностей измеренных X и Y. В дальнейшем будем говорить о нахождении только коэффициента регрессии K_1, так как K_0 вычисляется после нахождения K_1 по известной формуле

$$K_0 = \overline{Y} - K_1\overline{X}, \qquad (3)$$

где $\overline{X}$ и $\overline{Y}$ – средние значения X и Y.

2. Новый подход

Новый подход к нахождению коэффициента регрессии K_1 заключается в следующих двух моментах.

1. Предлагается X и Y нормировать соответственно на значения $\sqrt{\delta_X^2 + \delta_{X_0}^2}$ и $\sqrt{\delta_Y^2 + \delta_{Y_0}^2}$.

2. При нахождения K_1 используется ортогональная среднеквадратическая регрессия, т.е. минимизируется сумма квадратов отклонений, перпендикулярных искомой прямой.

Величины δ_X и δ_Y – случайные среднеквадратические погрешности измерения X и Y для рассматриваемого массива данных; δ_{X_0} и δ_{Y_0} – некоторые величины, характеризующие разброс точек в корреляционной связи физических величин X_0 и Y_0 за счет неконтролируемых физических параметров. Тогда уравнение линейной регрессии запишется в виде

$$\frac{Y}{\sqrt{\delta_Y^2+\delta_{Y_0}^2}} = K_0' + K_1' \frac{X}{\sqrt{\delta_X^2+\delta_{X_0}^2}}. \qquad (4)$$

Здесь величины δ_{X_0} и δ_{Y_0} находятся из решения системы двух уравнений.

Для получения первого уравнения воспользуемся известной формулой [1]:

$$\rho_{XY}\sigma_X\sigma_Y = \rho_{X_0Y_0}\sigma_{X_0}\sigma_{Y_0}, \qquad (5)$$

где ρ_{XY} — коэффициент корреляции между X и Y; σ_X и σ_Y — среднеквадратические отклонения X и Y; $\rho_{X_0Y_0}$ — коэффициент корреляции между X_0 и Y_0; $\sigma_{X_0}\sigma_{Y_0}$ — среднеквадратические отклонения X_0 и Y_0.

Тогда первое уравнение запишем в виде

$$\left|\rho_{X_0Y_0}\right|\sigma_{X_0}\sigma_{Y_0} = \sqrt{\sigma_{X_0}^2-\delta_{X_0}^2}\,\sqrt{\sigma_{Y_0}^2-\delta_{Y_0}^2}, \qquad (6)$$

где

$$\sigma_{X_0} = \sqrt{\sigma_X^2-\delta_X^2};\quad \sigma_{Y_0} = \sqrt{\sigma_Y^2-\delta_Y^2}.$$

Второе уравнение запишем в виде

$$\delta_{X_0}/\sigma_{X_0} = \delta_{Y_0}/\sigma_{Y_0} \qquad (7)$$

и назовем условием пропорциональности величин δ_{X_0}, δ_{Y_0} и σ_{X_0}, σ_{Y_0}. Введение величин δ_{X_0}, δ_{Y_0} и запись условия (7) являются ключевыми моментами в данной работе, так как это позволило получить обобщенное решение для коэффициентов линейной регрессии уравнения (2).

3. Результаты

После решения системы уравнений (6) и (7) получим

$$\delta_{X_0} = \sigma_X\sqrt{\left(1-\frac{\delta_X^2}{\sigma_X^2}\right)\left(1-\frac{\left|\rho_{XY}\right|}{\sqrt{\left(1-\delta_X^2/\sigma_X^2\right)\left(1-\delta_Y^2/\sigma_Y^2\right)}}\right)}, \qquad (8)$$

$$\delta_{Y_0} = \sigma_Y\sqrt{\left(1-\frac{\delta_Y^2}{\sigma_Y^2}\right)\left(1-\frac{\left|\rho_{XY}\right|}{\sqrt{\left(1-\delta_X^2/\sigma_X^2\right)\left(1-\delta_Y^2/\sigma_Y^2\right)}}\right)}. \qquad (9)$$

С учетом (8) и (9) найдем значения $\sqrt{\delta_X^2+\delta_{X_0}^2}$ и $\sqrt{\delta_Y^2+\delta_{Y_0}^2}$ в следующем виде:

$$\sqrt{\delta_X^2+\delta_{X_0}^2} = \sigma_X A, \qquad (10)$$

$$\sqrt{\delta_Y^2+\delta_{Y_0}^2} = \sigma_Y B, \qquad (11)$$

где

$$A = \sqrt{1-\left|\rho_{X_0Y_0}\right|\left(1-\frac{\delta_X^2}{\sigma_X^2}\right)} = \sqrt{1-\left|\rho_{XY}\right|\sqrt{\frac{1-\delta_X^2/\sigma_X^2}{1-\delta_Y^2/\sigma_Y^2}}}, \qquad (12)$$

$$B = \sqrt{1-\left|\rho_{X_0Y_0}\right|\left(1-\frac{\delta_Y^2}{\sigma_Y^2}\right)} = \sqrt{1-\left|\rho_{XY}\right|\sqrt{\frac{1-\delta_Y^2/\sigma_Y^2}{1-\delta_X^2/\sigma_X^2}}}. \qquad (13)$$

С учетом (10) и (11) уравнение линейной регрессии (4) запишется в виде

$$\frac{Y}{\sigma_Y B} = K_0' + K_1' \frac{X}{\sigma_X A}. \qquad (14)$$

Уравнение (14) легко привести к виду (2):

$$Y = K_0'\sigma_Y B + K_1'\frac{\sigma_Y B}{\sigma_X A}X = K_0 + K_1 X, \qquad (15)$$

где

$$K_0 = K_0' A\sigma_Y B, \qquad (16)$$

$$K_1 = K_1'\frac{\sigma_Y B}{\sigma_X A}. \qquad (17)$$

Применяя ортогональную среднеквадратическую регрессию к уравнению (14) и используя соотношение (17), получим выражение для искомого коэффициента регрессии:

$$K_1 = \frac{\sigma_Y}{\sigma_X}\frac{B}{A}\frac{1}{2\rho_{XY}}\left\{\left(\frac{A}{B}-\frac{B}{A}\right)+\sqrt{\left(\frac{A}{B}-\frac{B}{A}\right)^2+4\rho_{XY}^2}\right\}, \qquad (18)$$

где A и B определяются выражениями (12) и (13). Впервые формула (18) была представлена в [4], а подробно описана в [5].

4. Анализ

Выражение (18) позволяет устанавливать однозначную связь между X и Y и определять условия использования известных типов линейной регрессии.

Покажем, что все известные аналитические выражения для коэффициента регрессии K_1 уравнения (2) являются частными случаями формулы (18).

4.1. Так, для случая, когда разброс точек в корреляционной связи X и Y обусловлен только их случайными погрешностями, т.е. $\rho_{X_0Y_0} = 1$, получим известное выражение для коэффициента регрессии K_1, приведенное в [1]:

$$K_1 = \frac{\delta_Y}{\delta_X}\frac{1}{2\rho_{XY}}\left\{\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)+\right.$$

$$\left.+\sqrt{\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)^2+4\rho_{XY}^2}\right\}. \qquad (19)$$

4.1.1. При $\delta_X = 0$ и $\delta_Y \neq 0$ имеем

$$K_1 = \lim_{\delta_X\to 0}\frac{\delta_Y}{\delta_X}\frac{1}{2\rho_{XY}}\left\{\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)+\right.$$

$$+\left(-\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}+\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)\sqrt{1+4\rho_{XY}^2\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}\right)^2}\Bigg\}.$$

Разлагая выражение под квадратным корнем в ряд Маклорена [6] и оставляя первые два члена, получим

$$K_1=\lim_{\delta_X\to 0}\frac{\delta_Y}{\delta_X}\frac{1}{2\rho_{XY}}\times$$

$$\times\left\{\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)+\left(-\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}+\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)\times\right.$$

$$\left.\times\left[1+2\rho_{XY}^2\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}\right)^2\right]\right\}=\frac{\sigma_Y}{\sigma_X}\rho_{XY}. \quad (20)$$

Это известная формула для коэффициента K_1 уравнения прямой регрессии $Y = K_0 + K_1 X$, которая находится путем минимизации суммы квадратов отклонений вдоль оси Y от искомой прямой [2].

4.1.2. При $\delta_Y = 0$ и $\delta_X \neq 0$ имеем

$$K_1=\lim_{\delta_Y\to 0}\frac{\delta_Y}{\delta_X}\frac{1}{2\rho_{XY}}\left\{\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)+\right.$$

$$\left.+\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)\sqrt{1+4\rho_{XY}^2\left(\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)^2}\right\}.$$

Проведя процедуру разложения выражения под квадратным корнем в ряд Маклорена [6] и оставляя первые два члена, получим

$$K_1=\lim_{\delta_Y\to 0}\frac{\delta_Y}{\delta_X}\frac{1}{2\rho_{XY}}\left\{\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)+\right.$$

$$\left.+\left(\frac{\sigma_Y}{\sigma_X}\frac{\delta_X}{\delta_Y}-\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)\left[1+2\rho_{XY}^2\left(\frac{\sigma_X}{\sigma_Y}\frac{\delta_Y}{\delta_X}\right)^2\right]\right\}=\frac{\sigma_Y}{\sigma_X}\frac{1}{\rho_{XY}}. \quad (21)$$

Формула (21) – также известная формула для коэффициента $1/K_1^*$ уравнения обратной регрессии $X = K_0^* + K_1^* Y$, которая получается путем минимизации суммы квадратов отклонений вдоль оси X от искомой прямой [2].

4.1.3. При $\delta_X = \delta_Y \neq 0$ получим известную формулу

$$K_1=\frac{1}{2\rho_{XY}}\left\{\left(\frac{\sigma_Y}{\sigma_X}-\frac{\sigma_X}{\sigma_Y}\right)+\sqrt{\left(\frac{\sigma_Y}{\sigma_X}-\frac{\sigma_X}{\sigma_Y}\right)^2+4\rho_{XY}^2}\right\} \quad (22)$$

для коэффициента K_1 уравнения ортогональной регрессии $Y = K_0 + K_1 X$, которая находится путем минимизации суммы квадратов отклонений, перпендикулярных искомой прямой [3].

4.2. Если $\delta_X/\sigma_X = \delta_Y/\sigma_Y$, то из выражения (18) вытекает простая формула для коэффициента регрессии

$$K_1 = \sigma_Y/\sigma_X. \quad (23)$$

Заметим, что формула (23) представляет собой среднее геометрическое формул (20) и (21).

5. Диапазон изменчивости коэффициента регрессии

Для случая, когда разброс точек в корреляционной связи величин X и Y обусловлен только их случайными погрешностями, т.е. $\rho_{X_0Y_0} = 1$, коэффициент регрессии будет изменяться в следующих пределах:

$$\frac{\sigma_Y}{\sigma_X}|\rho_{XY}| \le |K_1| \le \frac{\sigma_Y}{\sigma_X}\frac{1}{|\rho_{XY}|}, \quad (24)$$

а при $\rho_{X_0Y_0} < 1$

$$\frac{\sigma_Y}{\sigma_X}|\rho_{XY}| < |K_1| < \frac{\sigma_Y}{\sigma_X}\frac{1}{|\rho_{XY}|}. \quad (25)$$

Как видно из выражений (24), (25), коэффициенты для прямой и обратной регрессий принимают соответственно минимальное и максимальное значения.

6. Вычисление случайных погрешностей из экспериментальных данных

Формула (18) имеет ценность только тогда, когда известны случайные погрешности δ_X и δ_Y. И чем точнее определены погрешности, тем меньше будет ошибка в вычислении коэффициента регрессии K_1.

В ряде случаев значения δ_X и δ_Y могут быть найдены из экспериментальных данных. Так, для X и Y можно получить по два подобных массива. Тогда, используя выражение (5), найдем искомые погрешности

$$\delta_X = \sigma_X\sqrt{1-|\rho_{XX}|}, \quad (26)$$

$$\delta_Y = \sigma_Y\sqrt{1-|\rho_{YY}|}, \quad (27)$$

где ρ_{XX} и ρ_{YY} – нормированные коэффициенты автокорреляции для X и Y.

При отсутствии исходных данных для нахождения погрешностей по точным формулам (26), (27) можно воспользоваться выражениями для их приближенной оценки. Для этого выберем величины X и X', Y и Y', незначительно отличающиеся друг от друга.

Полагая в (5) $\delta_X = \delta_{X'}$, $\delta_Y = \delta_{Y'}$, $\rho_{X_0X_0'} = 1$, $\rho_{Y_0Y_0'} = 1$, получим верхние оценки для δ_X и δ_Y:

$$\delta_X = \sqrt{\frac{\sigma_X^2+\sigma_{X'}^2}{2}-\sqrt{\left(\frac{\sigma_X^2-\sigma_{X'}^2}{2}\right)^2+\rho_{XX'}^2\sigma_X^2\sigma_{X'}^2}}, \quad (28)$$

$$\delta_Y = \sqrt{\frac{\sigma_Y^2 + \sigma_{Y'}^2}{2} - \sqrt{(\frac{\sigma_Y^2 - \sigma_{Y'}^2}{2})^2 + \rho_{YY'}^2\sigma_Y^2\sigma_{Y'}^2}}. \quad (29)$$

Заметим, что формулы (26), (27) являются частными случаями формул (28), (29) при условии, что $\sigma_X = \sigma_{X'}$ и $\sigma_Y = \sigma_{Y'}$ соответственно.

Если известна одна из погрешностей (δ_X или δ_Y), а разброс точек в искомой зависимости обусловлен только случайными погрешностями (т.е. $\rho_{X_0Y_0} = 1$), то согласно (5) значения других погрешностей будут вычисляться по следующим формулам:

$$\delta_X = \sigma_X\sqrt{1 - \rho_{XY}^2 \frac{\sigma_Y^2}{\sigma_Y^2 - \delta_Y^2}} \quad (30)$$

или

$$\delta_Y = \sigma_Y\sqrt{1 - \rho_{XY}^2 \frac{\sigma_X^2}{\sigma_X^2 - \delta_X^2}}. \quad (31)$$

Если при этом одна из погрешностей (δ_X или δ_Y) равна нулю, то, как следует из (30) и (31), значения других погрешностей будут вычисляться по формулам

$$\delta_X = \sigma_X\sqrt{1 - \rho_{XY}^2} \quad (32)$$

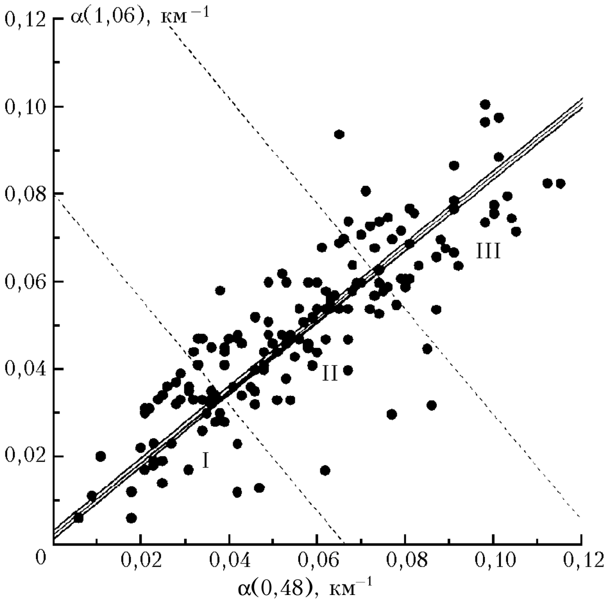
или

$$\delta_Y = \sigma_Y\sqrt{1 - \rho_{XY}^2}. \quad (33)$$

Таким образом, использование формул (26)–(33) для разных условий позволяет оценивать случайные погрешности измеряемых величин непосредственно из экспериментальных данных.

7. Применение обобщенной формулы для построения однопараметрической модели аэрозольного ослабления

Работоспособность и эффективность применения формулы (18) продемонстрируем на примере построения модели аэрозольного ослабления, позволяющей по измерениям коэффициентов аэрозольного ослабления в области спектра 0,48 мкм (X) рассчитывать аэрозольное ослабление в области 1,06 мкм (Y). Для этого воспользуемся экспериментальными данными, полученными в аридной зоне Казахстана в летний сезон [7]. На рисунке приведена взаимосвязь между коэффициентами аэрозольного ослабления на длинах волн 0,48 и 1,06 мкм. Разделим весь массив на три массива I, II и III, как показано на рисунке, пунктирными линиями и из их комбинаций сформируем четыре разных массива.



Для расчета случайных погрешностей δ_X и δ_Y воспользуемся формулами (28), (29), где в качестве недостающих величин X' и Y' возьмем массивы коэффициентов аэрозольного ослабления в ближайших областях спектра 0,55 и 0,87 мкм. Вычислив статистические характеристики коэффициентов аэрозольного ослабления на длинах волн 0,48, 0,55, 0,87 и 1,06 мкм для каждого массива, найдем значения случайных погрешностей.

В табл. 1 приведены номера сформированных массивов, их размерность, состав и значения статистических характеристик, которые необходимы для расчета коэффициентов регрессии линейного уравнения $\alpha(1,06) = K_0 + K_1\alpha(0,48)$ по обобщенной формуле (18) и по формулам (19), (20), (22), где $\alpha(1,06) = Y$, а $\alpha(0,48) = X$.

В табл. 2 приведены параметры моделей аэрозольного ослабления $\alpha(1,06) = K_0 + K_1\alpha(0,48)$, рассчитанные по формулам (18)–(20) и (22). Обращает на себя внимание тот факт, что модели, представляющие собой регрессию Y на X, получены для разных массивов по формуле (20) и характеризуются неустойчивыми значениями коэффициентов регрессии.

Таблица 1

Номер, размерность, состав и статистические характеристики сформированных массивов

Массив			ρ_{XY}	$\overline{X}$, км$^{-1}$	σ_X, км$^{-1}$	δ_X, км$^{-1}$	$\overline{Y}$, км$^{-1}$	σ_Y, км$^{-1}$	δ_Y, км$^{-1}$
№	размерность	состав							
1	160	I + II + III	0,84	0,057	0,0241	0,0063	0,049	0,0200	0,0050
2	120	I + II	0,71	0,047	0,0177	0,0058	0,041	0,0146	0,0051
3	120	II + III	0,72	0,067	0,0194	0,0066	0,057	0,0160	0,0049
4	80	II	0,33	0,056	0,0127	0,0071	0,049	0,0100	0,0049

Таблица 2

Параметры моделей аэрозольного ослабления $\alpha(1{,}06) = K_0 + K_1\alpha(0{,}48)$

№ массива	Формула							
	(20)		(22)		(19)		(18)	
	K_1	K_0, км$^{-1}$	K_1	K_0, км$^{-1}$	K_1	K_0, км$^{-1}$	K_1	K_0, км$^{-1}$
1	0,70	0,009	0,80	0,004	0,84	0,001	0,83	0,002
2	0,59	0,013	0,76	0,005	0,80	0,003	0,82	0,002
3	0,59	0,017	0,77	0,006	0,86	–0,001	0,83	0,001
4	0,26	0,034	0,50	0,020	1,00	–0,007	0,83	0,003

Модели, полученные по формуле (22) для ортогональной среднеквадратической регрессии и по формуле (19) для структурного соотношения [1], дают более стабильные значения коэффициентов регрессии, но также являются неустойчивыми, что хорошо видно на примере массива № 4.

В то же время модели, полученные для разных массивов по обобщенной формуле (18), мало отличаются друг от друга. Эти четыре модели показаны на рисунке сплошными линиями. Таким образом, формула (18) позволяет получать устойчивые и достоверные региональные модели, практически не зависящие от размерности массива и коэффициента корреляции ρ_{XY}. Кроме того, модели, получаемые по формуле (18), являются физически корректными. Так, аридная зона Казахстана в летний период характеризуется практически квазинейтральным спектральным ходом коэффициентов аэрозольного ослабления [7]. Поэтому коэффициент регрессии K_0 линейного уравнения $\alpha(1{,}06) = K_0 + K_1\alpha(0{,}48)$ должен быть близок к нулю. Как видно из табл. 2, обобщенная формула (18) дает наиболее близкие к нулю значения K_0 для всех четырех массивов. С учетом устойчивости получаемых коэффициентов регрессии и физической корректности результатов формула (18) является предпочтительной для построения линейных регрессионных моделей аэрозольного ослабления.

Заключение

Кратко сформулируем основные результаты.

1. Получена обобщенная формула, позволяющая находить коэффициенты регрессии линейного уравнения $Y = K_0 + K_1 X$ для общего случая, когда разброс точек в корреляционной связи величин X и Y обусловлен как их случайными погрешностями измерений, так и неконтролируемыми физическими факторами.

2. Все известные выражения для коэффициентов регрессии являются частными случаями полученной формулы.

3. Предложены способы и приведены выражения для оценки случайных среднеквадратических погрешностей измеряемых величин, входящих в формулу для вычисления коэффициентов регрессии, из экспериментальных данных.

4. Обобщенная формула позволяет получать устойчивые, достоверные и физически корректные однопараметрические модели.

Полученная формула представляет интерес для специалистов, занимающихся обработкой экспериментальных данных, и может быть использована для корректной физической интерпретации, независимо от области знания.

В заключение автор выражает свою признательность д.т.н. профессору ТУСУРа А.А. Мицелю за сделанные полезные замечания.

1. *Кендалл М., Стьюарт А.* Статистические выводы и связи. М.: Наука, 1973. Т. 2. 900 с.
2. *Зайдель А.Н.* Погрешности измерений физических величин. Л.: Наука, 1985. 112 с.
3. *Крамер Г.* Математические методы статистики. М.: Мир, 1975. 648 с.
4. *Щелканов Н.Н.* Построение регрессионной зависимости между аэрозольными оптическими толщами атмосферы с учетом их случайных погрешностей // II Заседание Рабочей группы проекта «Аэрозоли Сибири»: Тезисы докл. Томск: Изд-во ИОА СО РАН, 1995. С. 16.
5. *Щелканов Н.Н.* Обобщенный метод построения линейной регрессии между двумя физическими величинами с учетом их случайных погрешностей. Препр. / Ин-т оптики атмосферы СО РАН (Томск). 2003. № 2. 15 с.
6. *Кудрявцев В.А., Демидович Б.П.* Краткий курс высшей математики. М.: Наука, 1975. 624 с.
7. *Пхалагов Ю.А., Ужегов В.Н., Щелканов Н.Н.* Аэрозольное ослабление оптического излучения в атмосфере аридной зоны // Оптика атмосф. и океана. 1994. Т. 7. № 10. С. 1318–1329.

N.N. Shchelkanov. **Generalized method for construction of linear regression and its application to development of one-parameter aerosol extinction models.**

A generalized equation is presented for determination of the regression coefficients of the linear equation $Y = K_0 + K_1X$ in the general case, when the point spread in the correlation between X and Y is caused both by random measurement errors and by uncontrollable physical factors. All the known equations for the regression coefficients appeared to be particular cases of the equation obtained. Methods are proposed and equations are presented for estimation of rms errors of measured parameters, entering into the equation for calculation of the regression coefficients, from experimental data.