

# Регрессионный анализ

- линейная зависимость между переменными,
- метод наименьших квадратов и другие методы оценки коэффициентов,
- оценка эффективности коэффициентов уравнения регрессии
- простая и множественная регрессия,
- методы построения уравнений регрессии,
- методы анализа остатков.

ЗАРУБЕЖНЫЕ  
СТАТИСТИЧЕСКИЕ  
ИССЛЕДОВАНИЯ



---

**ПРИКЛАДНОЙ  
РЕГРЕССИОННЫЙ  
АНАЛИЗ**

•  
**Н. ДРЕЙПЕР  
Г. СМИТ**

APPLIED  
REGRESSION ANALYSIS

---

N. R. DRAPER, H. SMITH

JOHN WILEY AND SONS, INC.  
NEW YORK LONDON SYDNEY

ПРИКЛАДНОЙ  
РЕГРЕССИОННЫЙ АНАЛИЗ

---

Н. ДРЕЙПЕР, Г. СМИТ

Перевод с английского, научное редактирование  
и предисловие Ю. П. Адлера и В. Г. Горского

---

«СТАТИСТИКА» МОСКВА 1973

## СОДЕРЖАНИЕ

Предисловие к русскому изданию . . . . .	5
<b>Глава 1. Подбор прямой методом наименьших квадратов . . . . .</b>	<b>9</b>
1.0. Введение. Потребность в статистическом анализе . . . . .	9
1.1. Линейная зависимость между двумя переменными . . . . .	13
1.2. Линейная регрессия: подбор прямой . . . . .	15
1.3. Точность оценки регрессии . . . . .	22
1.4. Исследование уравнения регрессии . . . . .	25
1.5. Неадекватность и «чистая» ошибка . . . . .	34
1.6. Корреляция между $X$ и $Y$ . . . . .	40
Упражнения . . . . .	43
<b>Глава 2. Матричный подход к линейной регрессии . . . . .</b>	<b>53</b>
2.0. Введение . . . . .	53
2.1. Подбор уравнения прямой в матричных обозначениях; оценки параметров $\beta_0$ и $\beta_1$ . . . . .	53
2.2. Дисперсионный анализ в матричных обозначениях . . . . .	63
2.3. Вычисления дисперсий и ковариаций коэффициентов $b_0$ и $b_1$ в матричной форме . . . . .	64
2.4. Дисперсия величин $\bar{Y}$ в матричных обозначениях . . . . .	65
2.5. Резюме к матричному подходу при подборе прямой . . . . .	65
2.6. Случай общей регрессии . . . . .	67
2.7. Принцип «дополнительной суммы квадратов» . . . . .	76
2.8. Ортогональные столбцы в матрице $X$ . . . . .	78
2.9. Частный и последовательный $F$ -критерий . . . . .	80
2.10. Проверка общей линейной гипотезы в регрессионных задачах . . . . .	82
2.11. Взвешенный метод наименьших квадратов . . . . .	86
2.12. Смещение регрессионных оценок . . . . .	90
<b>Глава 3. Исследование остатков . . . . .</b>	<b>95</b>
3.0. Введение . . . . .	95
3.1. Общий график . . . . .	96
3.2. График временной последовательности . . . . .	98
3.3. График зависимости остатков от $\hat{Y}_i$ . . . . .	99
3.4. График зависимости остатков от независимых переменных $X_{ji}$ , $i = 1, 2, 3, \dots, n$ . . . . .	100
3.5. Другие графики остатков . . . . .	101
3.6. Статистики для исследования остатков . . . . .	102
3.7. Корреляция между остатками . . . . .	103
3.8. Выбросы . . . . .	104
3.9. Исследование серий на графиках временной последовательности остатков . . . . .	104
Упражнения . . . . .	109
<b>Глава 4. Две независимые переменные . . . . .</b>	<b>115</b>
4.0. Введение . . . . .	115
4.1. Сведение множественной регрессии с двумя независимыми переменными к последовательности однофакторных линейных регрессий . . . . .	118
4.2. Исследование уравнения регрессии . . . . .	125
Упражнения . . . . .	133

<b>Глава 5. Более сложные модели . . . . .</b>	<b>137</b>
5.0. Введение . . . . .	137
5.1. Полиномиальные модели различных порядков по $X_j$ . . . . .	138
5.2. Модели, включающие преобразования, отличные от целых степеней . . . . .	139
5.3. Использование «фиктивных» переменных в множественной регрессии . . . . .	143
5.4. Подготовка матрицы исходных данных для решения общей задачи регрессии . . . . .	150
5.5. Ортогональные полиномы . . . . .	159
5.6. Преобразование матрицы $X$ для получения ортогональных столбцов . . . . .	164
Упражнения . . . . .	167
<b>Глава 6. Выбор «наилучшего» уравнения регрессии . . . . .</b>	<b>172</b>
6.0. Введение . . . . .	172
6.1. Метод всех возможных регрессий . . . . .	173
6.2. Метод исключения . . . . .	177
6.3. Метод включения . . . . .	178
6.4. Шаговый регрессионный метод . . . . .	180
6.5. Две вариации четырех предыдущих методов . . . . .	182
6.6. Ступенчатый регрессионный метод . . . . .	183
6.7. Сводка МНК-уравнений, полученных описанными методами . . . . .	187
6.8. Вычислительные аспекты шагового регрессионного метода . . . . .	188
Упражнения . . . . .	203
<b>Глава 7. Типичный пример . . . . .</b>	<b>226</b>
7.0. Введение . . . . .	226
7.1. Задача . . . . .	226
7.2. Исследование данных . . . . .	226
7.3. Выбор первого фактора для включения в регрессию . . . . .	228
7.4. Построение новых переменных . . . . .	231
7.5. Включение в модель взаимодействия . . . . .	231
7.6. Расширение модели . . . . .	232
Упражнения . . . . .	234
<b>Глава 8. Множественная регрессия и построение математической модели . . . . .</b>	<b>242</b>
8.0. Введение . . . . .	242
8.1. Планирование процесса построения модели . . . . .	244
8.2. Разработка математической модели . . . . .	247
8.3. Проверка и использование математической модели . . . . .	248
<b>Глава 9. Приложение множественной регрессии к задачам дисперсионного анализа . . . . .</b>	<b>250</b>
9.0. Введение . . . . .	250
9.1. Односторонняя классификация . . . . .	251
9.2. Регрессионная обработка односторонней классификации с использованием исходной модели . . . . .	252
9.3. Регрессионная обработка односторонней классификации: независимые нормальные уравнения . . . . .	257
9.4. Двусторонняя классификация с равным числом наблюдений в ячейках . . . . .	259
9.5. Регрессионная обработка двусторонней классификации с равным числом наблюдений в ячейках . . . . .	260
9.6. Пример: двусторонняя классификация . . . . .	264
9.7. Комментарии . . . . .	266
Упражнения . . . . .	266

Глава 10. Введение в нелинейное оценивание . . . . .	269
10.1. Введение . . . . .	269
10.2. Метод наименьших квадратов в нелинейном случае . . . . .	270
10.3. Оценивание параметров нелинейных систем . . . . .	273
10.4. Пример . . . . .	282
10.5. Некоторые замечания о репараметризации модели . . . . .	290
10.6. Геометрия линейного метода наименьших квадратов . . . . .	292
10.7. Геометрия нелинейного метода наименьших квадратов . . . . .	302
Упражнения . . . . .	306
Библиография по нелинейному регрессионному анализу . . . . .	308
Процентные точки $t$ -распределения . . . . .	312
Процентные точки $F$ -распределения . . . . .	313
Библиография . . . . .	315
Ответы к упражнениям . . . . .	322
Приложение А . . . . .	346
Приложение Б . . . . .	357
Указатели . . . . .	377

**Химмельблау Д. Прикладное нелинейное программирование. М. – Мир., 1975. – 534 с.**

Государственный ордена Трудового Красного Знамени  
гидрологический институт

на правах рукописи

ЛОБАНОВ Владимир Алексеевич

УДК 556.048

СТАТИСТИЧЕСКИЕ КРИТЕРИИ И РЕГРЕССИЯ  
ПРИ АНАЛИЗЕ МНОГОЛЕТНИХ КОЛЕБАНИЙ РЕЧНОГО СТОКА

05.14.09 – Гидравлика и инженерная гидрология

Д и с с е р т а ц и я  
на соискание ученой степени  
кандидата технических наук

Научный руководитель:  
доктор технических наук  
А.В.Рожественский

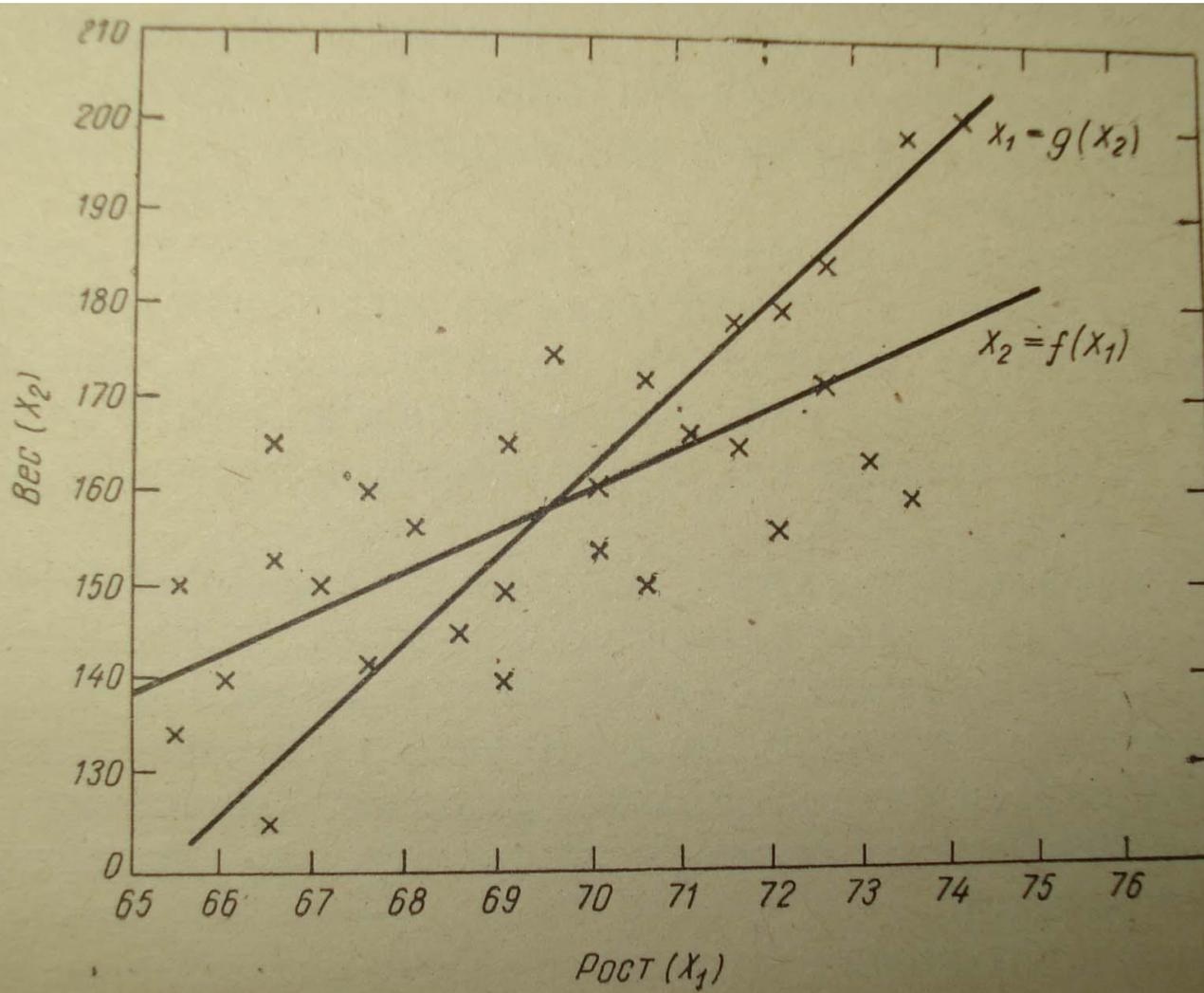
Ленинград – 1983

ВВЕДЕНИЕ . . . . .	5
I. ОБОБЩЕНИЕ КРИТЕРИЕВ ОДНОРОДНОСТИ, СЛУЧАЙНОСТИ И СОГЛАСИЯ ДЛЯ РАСПРЕДЕЛЕНИЯ ПИРСОНА III ТИПА И МОДЕЛИ ПРОСТОЙ ЦЕПИ МАРКОВА. . . . .	10
I.1. Основные предпосылки методов математической статисти- стики и особенности рядов гидрологических характе- ристик. . . . .	10
I.2. Алгоритмы расчета статистик критериев с использова- нием метода статистических испытаний. . . . .	13
I.3. Обобщение критериев оценки резко отклоняющихся элементов выборки для распределения Пирсона III типа и модели простой цепи Маркова . . . . .	16
I.3.1. Выбор критериев и их свойства. . . . .	16
I.3.2. Обобщение критериев Диксона. . . . .	18
I.3.3. Обобщения критерия Смирнова-Граббса. . . . .	32
I.3.4. Оценка эффективности обобщенных критериев Диксона и Смирнова-Граббса . . . . .	41
I.3.5. Оценка однородности эмпирической функции распределения с использованием обобщенных критериев Диксона и Смирнова-Граббса . . . . .	51
I.3.6. Примеры использования обобщенных критериев Диксона и Смирнова-Граббса . . . . .	54
I.4. Некоторые вопросы анализа стационарности временных рядов . . . . .	61
I.5. Анализ критериев случайности временных рядов. . . . .	66
I.6. Обобщение критериев согласия для модели простой цепи Маркова. . . . .	76
2. ОЦЕНКА ОДНОРОДНОСТИ, СТАЦИОНАРНОСТИ И СЛУЧАЙНОСТИ	

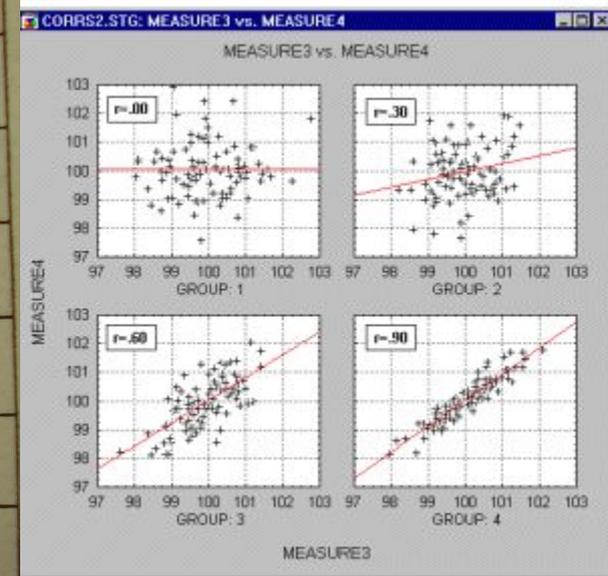
	Стр.
МНОГОЛЕТНИХ КОЛЕБАНИЙ РЕЧНОГО СТОКА . . . . .	85
2.1. Используемые данные наблюдений . . . . .	85
2.2. Анализ многолетних колебаний максимальных расходов воды . . . . .	94
2.3. Анализ многолетних колебаний минимальных летних расходов воды . . . . .	106
2.4. Анализ многолетних колебаний среднегодовых расхо- дов воды . . . . .	112
2.4.1. Анализ многолетних колебаний среднегодовых расходов воды рек ЕТС . . . . .	112
2.4.2. Анализ многолетних колебаний среднегодовых расходов воды в пределах отдельных районов . . . . .	116
2.5. Анализ многолетних колебаний объемов стока весен- него половодья . . . . .	120
2.6. Анализ многолетних колебаний факторов речного стока . . . . .	121
3. ОБОБЩЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ КОЭФФИЦИЕНТОВ УРАВНЕ- НИЯ РЕГРЕССИИ И КОЭФФИЦИЕНТОВ МНОЖЕСТВЕННОЙ КОРРЕЛЯЦИИ ДЛЯ РАСПРЕДЕЛЕНИЯ ПИРСОНА III ТИПА И МОДЕЛИ ПРОСТОЙ ЦЕПИ МАРКОВА . . . . .	126
3.1. Общие принципы регрессионного анализа и особенно- сти его применения в гидрологии . . . . .	126
3.2. Выбор и описание алгоритма статистического модели- рования . . . . .	132
3.3. Влияние асимметрии и автокорреляции на распределе- ния выборочных коэффициентов множественной корре- ляции . . . . .	138
3.4. Влияние асимметрии и автокорреляции на распределе- ния выборочных коэффициентов уравнения парной ре- грессии . . . . .	155

3.5. Влияние асимметрии и автокорреляции на распределе- ния выборочных свободных членов уравнения парной регрессии . . . . .	162
3.6. Обобщение полученных доверительных интервалов для коэффициентов уравнения множественной линейной регрессии . . . . .	168
3.7. Применение регрессионного анализа для построения уравнений формирования объема стока весеннего половодья . . . . .	176
3.8. Применение регрессионного анализа для оптимизации коэффициентов уравнений водохозяйственных балансов . . . . .	188
4. НЕКОТОРЫЕ ВОПРОСЫ ПРИМЕНЕНИЯ РЕГРЕССИОННОГО АНАЛИЗА В РАСЧЕТАХ СТОКА . . . . .	201
4.1. Особенности гидрологических расчетов . . . . .	201
4.2. Оценка эквивалентных периодов среднего и коэффици- ента вариации при приведении коротких рядов наблю- дений за стоком к многолетнему периоду . . . . .	203
4.3. Оценка эквивалентных периодов квантилей при приве- дении рядов наблюдений за стоком к многолетнему периоду по уравнению регрессии . . . . .	215
4.4. Применение метода статистических испытаний для расчета стока по композиции распределений стоко- формирующих факторов . . . . .	225
4.5. Расчет доверительных интервалов случайных ошибок для квантилей распределения Пирсона III типа . . . . .	231
ЗАКЛЮЧЕНИЕ . . . . .	238
ЛИТЕРАТУРА . . . . .	242
ПРИЛОЖЕНИЕ . . . . .	254

# Линейная зависимость между переменными

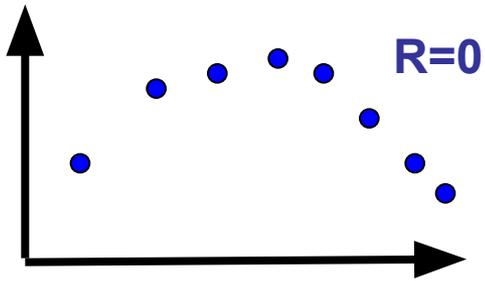


$$r = \frac{\sum (x_{1i} - \bar{x}_1) \cdot (x_{2i} - \bar{x}_2)}{\sqrt{\sum (x_{1i} - \bar{x}_1)^2} \cdot \sqrt{\sum (x_{2i} - \bar{x}_2)^2}}$$

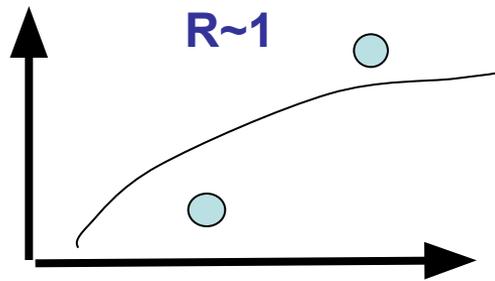


$$Y = b_1 X + b_0 \pm \varepsilon$$

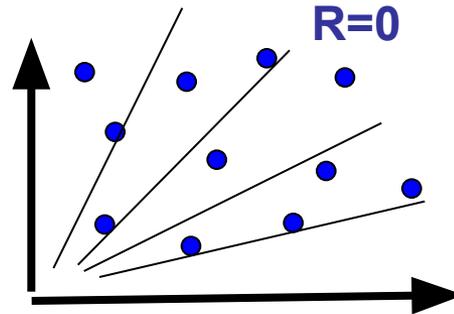
# Линейная зависимость между переменными



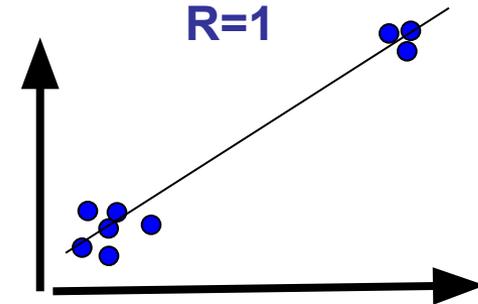
Функциональные преобразования факторов



Статистическая  
незначимость

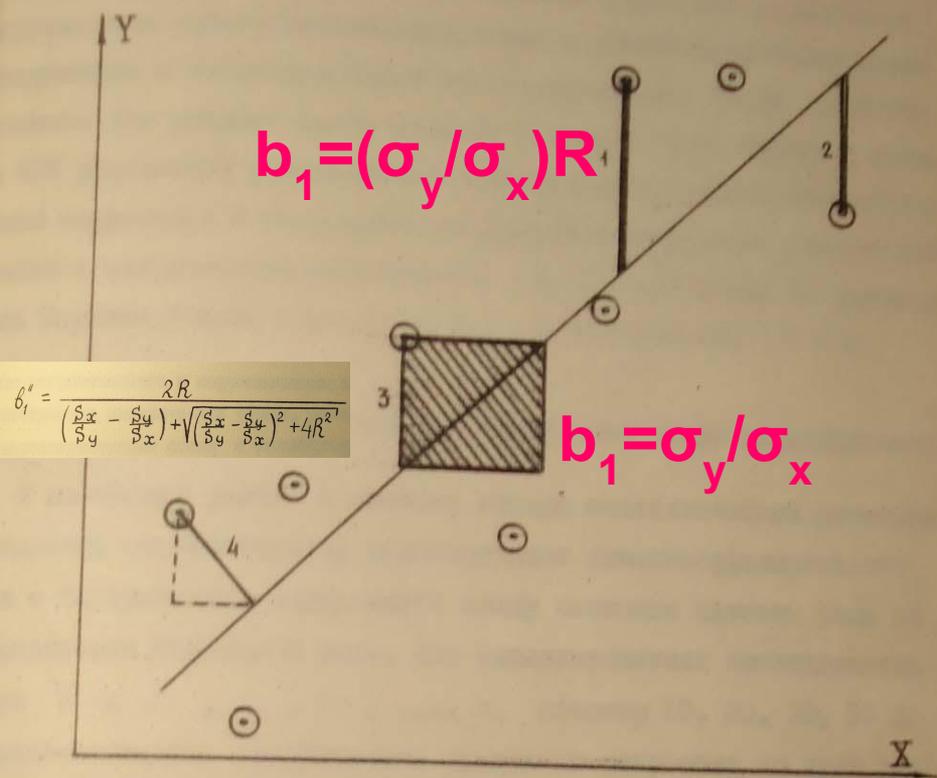


Совместный эффект



Ложная корреляция

# Метод наименьших квадратов и другие методы оценки коэффициентов



$$b_1 = (\sigma_y / \sigma_x) R$$

$$b_1 = \sigma_y / \sigma_x$$

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

Рис. 11. Геометрическая интерпретация методов расчета коэффициентов уравнения регрессии:

- 1 - МНК,  $\min \sum \Delta Y_i^2$ .
- 2 - метод абсолютных отклонений,  $\min \sum |\Delta Y_i|$ .
- 3 - метод однозначной аппроксимации,  $\min \sum \Delta Y_i \Delta X_i$ .
- 4 - метод ортогональной регрессии,  $\min \sum \sqrt{\Delta Y_i^2 + \Delta X_i^2}$ .

## МНК

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (1.2.4)$$

Будем подбирать значения оценок  $b_0$  и  $b_1$  так, чтобы их подстановка вместо  $\beta_0$  и  $\beta_1$  в уравнение (1.2.4) давала наименьшее возможное (минимальное) значение  $S$ . (Заметим, что  $X_i, Y_i$  — это фиксированные числа, которые нам известны.) Мы можем определить  $b_0$  и  $b_1$  путем дифференцирования уравнения (1.2.4) сначала по  $b_0$ , затем по  $b_1$  и приравнивания результатов к нулю. Тогда

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i),$$

$$\frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^n X_i (Y_i - b_0 - b_1 X_i), \quad (1.2.5)$$

так что для оценок  $b_0$  и  $b_1$  имеем

$$\sum_{i=1}^n (Y_i - b_0 - b_1 X_i) = 0,$$

$$\sum_{i=1}^n X_i (Y_i - b_0 - b_1 X_i) = 0, \quad (1.2.6)$$

где при приравнивании выражений (1.2.5) к нулю мы подставили  $(b_0, b_1)$  вместо  $(\beta_0, \beta_1)$ . Из (1.2.6) имеем:

$$\sum_{i=1}^n Y_i - n b_0 - b_1 \sum_{i=1}^n X_i = 0,$$

$$\sum_{i=1}^n X_i Y_i - b_0 \sum_{i=1}^n X_i - b_1 \sum_{i=1}^n X_i^2 = 0 \quad (1.2.7)$$

или

$$b_0 n + b_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i,$$

$$b_0 \sum_{i=1}^n X_i + b_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i. \quad (1.2.8)$$

Эти уравнения называют *нормальными*.

Решение уравнений (1.2.8) относительно  $b_1$  дает

$$b_1 = \frac{\sum X_i Y_i - [(\sum X_i)(\sum Y_i)]/n}{\sum X_i^2 - (\sum X_i)^2/n} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \quad (1.2.9)$$

## b\_0 = Y\_ср - b\_1 X\_ср

# Оценка эффективности коэффициентов уравнения регрессии

## Коэффициент $b_1$ уравнения

та.) Стандартная ошибка  $b_1$  есть корень квадратный из дисперсии, т. е.

$$\text{ст. ош. } (b_1) = \frac{\sigma}{\{\sum (X_i - \bar{X})^2\}^{1/2}}.$$

Если  $\sigma$  неизвестна и мы применяем вместо нее оценку  $s$ , предполагая, что модель корректна, то оценка стандартной ошибки  $b_1$  есть

$$\text{оц. ст. ош. } (b_1) = \frac{s}{\{\sum (X_i - \bar{X})^2\}^{1/2}}. \quad (1.4.2)$$

Если мы предполагаем, что разброс наблюдений относительно линии нормален, т. е., что ошибки  $\varepsilon_i$  все принадлежат некоторому нормальному распределению,  $N(0, \sigma^2)$ , то можно показать, что 100  $(1 - \alpha)$ %-ные доверительные интервалы для  $\beta_1$  получатся, если вычислить

$$b_1 \pm \frac{t\left(n-2, 1-\frac{1}{2}\alpha\right) s}{\{\sum (X_i - \bar{X})^2\}^{1/2}}, \quad (1.4.3)$$

## Коэффициент $b_0$ уравнения

$$\text{кв. ош. } (b_0) = \left\{ \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} \right\}^{1/2} \sigma.$$

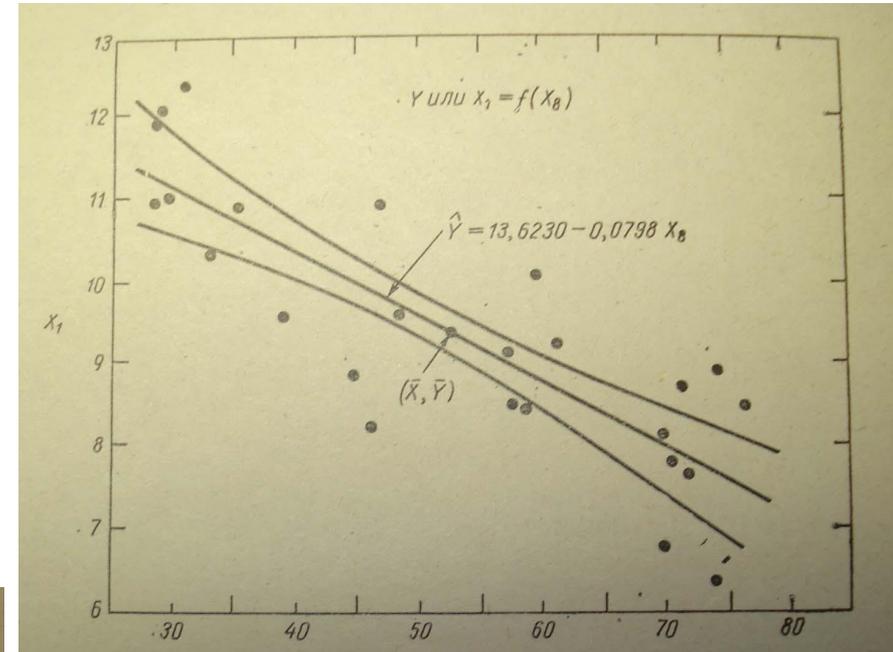
Отсюда имеем 100  $(1 - \alpha)$ %-ные доверительные пределы для  $\beta_0$

$$b_0 \pm t\left(n-2, 1-\frac{1}{2}\alpha\right) \left\{ \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} \right\}^{1/2} s.$$

## Расчетное значение $\hat{Y}$

$$\text{оц. кв. ош. } (\hat{Y}_k) = s \left\{ \frac{1}{n} + \frac{(X_k - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right\}^{1/2}.$$

## Доверительные интервалы уравнения регрессии



# Простая и множественная регрессия

$$Y = b_1 X + b_0 \pm \varepsilon$$

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_0 \pm \varepsilon$$

аддитивная

$$Y = b_1 * X_1 * X_2 + \dots + b_0 \pm \varepsilon$$

мультипликативная

$$Y = b_1 X + b_2 * X_1 * X_2 + \dots + b_0 \pm \varepsilon$$

смешанная с совместными эффектами

$$Y = b_1 X^2 + b_2 * X_1 * X_2 + b_3 \log X_3 + \dots + b_0 \pm \varepsilon$$

смешанная с совместными эффектами и функциональными преобразованиями факторов

# Методы построения уравнений регрессии

## 1. Метод исключения

Алгоритм:

- рассчитывается регрессионное уравнение, включающее все переменные (факторы);
- рассчитывается величина частного  $F$ -критерия для каждой из рассматриваемых переменных, как будто бы она была последней переменной, введенной в регрессионное уравнение;
- наименьшая величина частного  $F$ -критерия, обозначаемая через  $FL$ , сравнивается с заранее выбранным уровнем значимости ( $F_0$ );
- если  $FL < F_0$ , то переменная  $X_L$ , связанная с  $FL$ , исключается из рассмотрения и производится перерасчет уравнения регрессии с учетом оставшихся переменных, затем снова рассчитывается величина частного  $F$ -критерия для каждой из оставшихся переменных и процедура повторяется;
- если  $FL > F_0$ , то остается то уравнение, которое построено.

$$F_{\text{част } x_i} = \frac{R^2_{y, x_1 \dots x_i \dots x_p} - R^2_{y, x_1 \dots x_{i-1} x_{i+1} \dots x_p}}{1 - R^2_{y, x_1 \dots x_i \dots x_p}} \cdot \frac{n - m - 1}{1}$$

Частный  $F$ -критерий - отношение среднего квадрата новой переменной  $X_i$  с одной степенью свободы к дисперсии модели.

# Методы построения уравнений регрессии

## 2. Шаговая процедура

Включаются все переменные по очереди до тех пор, пока регрессионное уравнение не станет удовлетворительным.

Алгоритм:

- прежде всего выбирается переменная  $X_1$ , имеющая наиболее высокий коэффициент парной корреляции с  $Y$  и определяется расчетное значение  $Y_p$  по однофакторному уравнению с  $X_1$ ;
- определяется частный коэффициент корреляции между остатками  $E=Y-Y_p$  и остальными переменными за исключением  $X_1$ ;
- выбирается величина  $X_2$ , которая имеет наибольший частный коэффициент корреляции и находится второе регрессионное уравнение  $Y=f(X_1, X_2)$ ;
- далее включенная переменная  $X_2$  исследуется на эффективность включения по частному  $F$ -критерию таким же образом, как и в методе исключения;
- также, если  $FL > F_0$ , то переменная остается в уравнении, если  $FL < F_0$ , то исключается и процедура повторяется для следующей переменной с наибольшим частным коэффициентом корреляции.

# Оценка эффективности уравнения

## Фундаментальное уравнение регрессионного анализа

Уравнение (1.3.2) можно записать иначе<sup>12</sup>:

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2. \quad (1.3.3)$$

Теперь  $Y_i - \bar{Y}$  — это отклонение  $i$ -го наблюдения от общего среднего, следовательно, левая часть уравнения (1.3.3) — это сумма квадратов отклонений наблюдений от среднего; сокращенно — *SS относительно среднего*, а также *скорректированная сумма квадратов Y-ов*. Так как  $Y_i - \hat{Y}_i$  есть отклонение  $i$ -го наблюдения от его предсказанного или вычисленного значения ( $i$ -й остаток — см. гл. 3), а  $\hat{Y}_i - \bar{Y}$  есть отклонение предсказанного значения  $i$ -го наблюдения от среднего, то мы можем выразить уравнение (1.3.3) следующим образом<sup>12</sup>:

$$\left( \begin{array}{c} \text{Сумма квадратов} \\ \text{относительно среднего} \end{array} \right) = \left( \begin{array}{c} \text{Сумма квадратов} \\ \text{относительно регрес-} \\ \text{сии} \end{array} \right) + \left( \begin{array}{c} \text{Сумма квадратов, обу-} \\ \text{словленная регрессией} \end{array} \right)^{12A}$$

<sup>12</sup> Уравнение (1.3.3) играет фундаментальную роль в дисперсионном анализе. Можно даже сказать, что в нем в зародыше содержится весь дисперсионный анализ. — Прим. перев.

## Таблица дисперсионного анализа

Источник	Сумма квадратов	Степень свободы	Средний квадрат
Регрессия	$b_1 \left\{ \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n} \right\}$	1	$MS_R$
Относительно регрессии (остаток).	По разности	$n - 2$	$s^2 = \frac{(SS)}{(n-2)}$
Относительно среднего (общий, скорректированный на среднее)	$\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$	$n - 1$	

# Методы анализа остатков

Остатки любой эмпирической зависимости определяются как разности между фактическими (наблюдеными) и расчетными значениями:

$$\varepsilon_i = Y_i - \hat{Y}_i, \quad \text{или} \quad \Delta_i = (Y_i - \hat{Y}_i) / Y_i,$$

где:  $Y_i$  - фактическое (наблюденное) значение,  $\hat{Y}_i$  - рассчитанное по зависимости,  $\varepsilon_i$  – остаток или погрешность полученной зависимости.

Наиболее распространенной обобщенной характеристикой остатков является их среднее квадратическое отклонение ( $\sigma_i$ ):

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{n-1}} \quad \text{или:} \quad \sigma_\varepsilon = \sigma_Y \sqrt{1 - R^2}$$

В качестве обобщенной меры может служить также величина:

$$\Delta' = (1 - R^2) * 100\%,$$

которая характеризует долю исходного рассеивания (в %), не объясненного с помощью построенной зависимости.

# Методы анализа остатков

## Всесторонний анализ остатков включает в себя оценивание:

- резко отклоняющихся экстремальных значений;
- смещенности остатков;
- случайности остатков на основе хронологического графика;
- случайности остатков в зависимости от каждого фактора, входящего в уравнение;
- случайности остатков в зависимости от расчетных значений.

## Смещенность

Наличие смещенности остатков определяется тем, что среднее их значение не равно нулю или статистически значимо отличается от нуля.

Как правило, наличие смещенности может иметь место для уравнений балансового вида (уравнение водного, руслового и других видов баланса), где невязки уравнений характеризуют как неучтенные факторы, так и все систематические погрешности составляющих.

Смещенность остатков необходимо исключать или путем корректировки свободного члена уравнения или тех коэффициентов и факторов, которые ее обусловили.

# Методы анализа остатков

## Случайность остатков во времени

### Два пути:

- применение известных статистических критериев оценки случайности и стационарности (критерии Стьюдента, Фишера и другие);
- применение графического анализа остатков в зависимости от времени.

### При неслучайном характере возможны следующие основные варианты:

- полоса разброса остатков сужается или расширяется, что связано с непостоянством дисперсии остатков во времени;
- полоса остатков имеет одинаковую ширину, но изменяется (линейно или нелинейно) в зависимости от времени, что свидетельствует о нестационарности средних значений остатков.

# Случайность остатков от факторов и расчетного значения

## Возможны следующие ситуации:

- зависимость отсутствует и полоса остатков горизонтальна и симметрична относительно нулевого значения, что свидетельствует о случайности погрешностей;
- зависимость представлена сужающейся или расширяющейся полосой остатков от фактора, что свидетельствует о неоднородности дисперсии остатков, которую надо учитывать взвешенным МНК или предварительным преобразованием  $Y_i$ ;
- линейная зависимость остатков от фактора свидетельствует о том, что линейный эффект данного фактора в уравнении исключен неверно;
- нелинейная зависимость остатков от фактора свидетельствует о том, что в уравнение необходимо ввести нелинейные члены от  $X_i$  или произвести преобразование  $Y_i$ .

## Проверка построенного эмпирического уравнения на независимом от расчета материале наблюдений.

Анализ остатков в случае независимой проверки осуществляется теми же способами: на резко отклоняющиеся экстремумы, в зависимости от времени, факторов и расчетного значения.

Необходимо отметить, что должен иметь место оптимум между количеством информации, используемой для построения зависимости и для ее независимой проверки.

# Методы анализа остатков

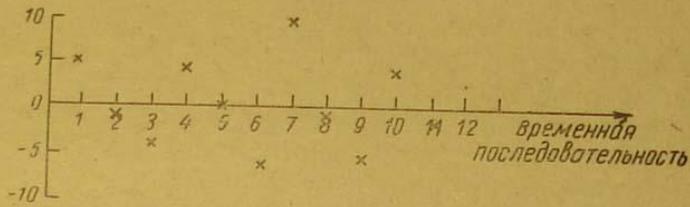


Рис. 3.2.



Рис. 3.3.

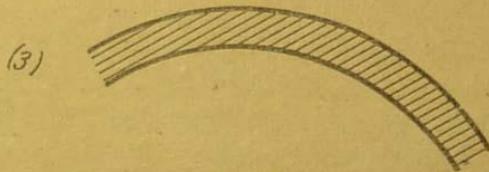
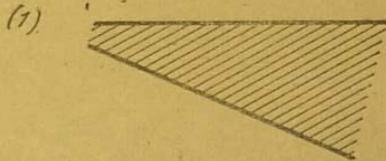


Рис. 3.4.

(1) Дисперсия не постоянна, а расчет со временем – следует использовать взвешенный МНК

(2) В модель следует включить фактор времени

(3) В модель должны быть включены линейный и квадратичный члены от времени.

## **ПРИМЕР** Оценка эффективности эмпирической зависимости

Слой поверхностного стока весеннего половодья (Y) на р.Оке – с.Половское

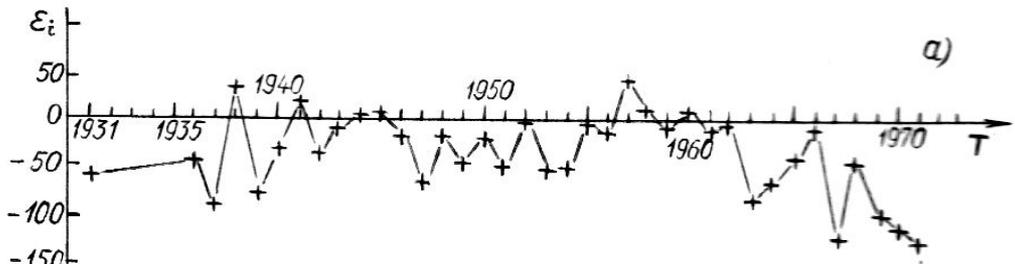
$$Y = 0.80X_1 + 0.86X_2 - 104,$$

где:  $X_1 = S + S_{л} + X_{ос}$ ,

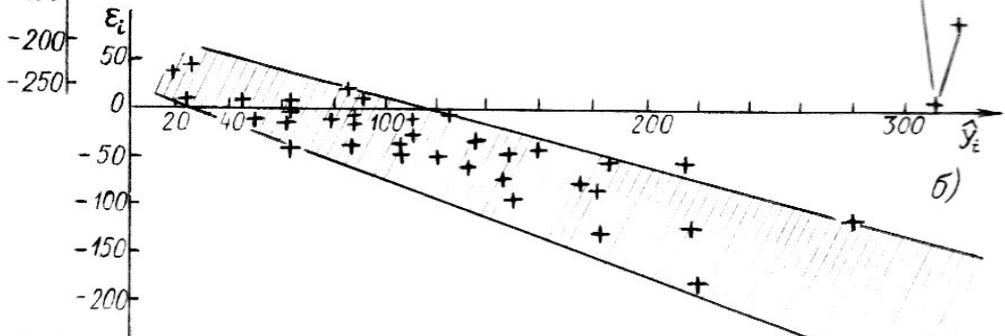
S - максимальные запасы воды в снеге (мм),  $S_{л}$  – запас воды в ледяной корке (мм),  $X_{ос}$  – осадки за период половодья (мм);

$X_2 = (L * e)/50$ , L – глубина промерзания почвы (см), e – величина осеннего увлажнения почвы (см).

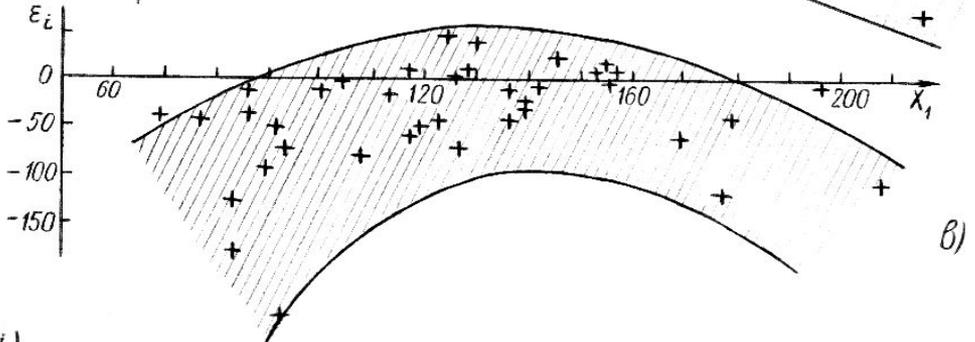
# Анализ остатков эмпирической зависимости для расчета слоев половодья



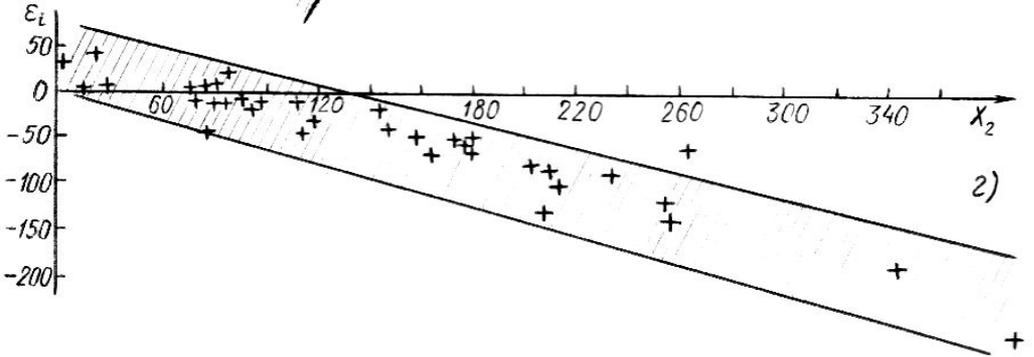
а) ЗАВИСИМОСТЬ ОСТАТКОВ ОТ ВРЕМЕНИ ( $\epsilon=f(t)$ )



б) ЗАВИСИМОСТЬ ОСТАТКОВ ОТ РАСЧЕТНЫХ СЛОЕВ СТОКА ( $\epsilon=f(Y')$ )



в) ЗАВИСИМОСТЬ ОСТАТКОВ ОТ ПЕРВОГО ФАКТОРА ( $\epsilon=f(X_1)$ )



г) ЗАВИСИМОСТЬ ОСТАТКОВ ОТ ВТОРОГО ФАКТОРА ( $\epsilon=f(X_2)$ )

## Выводы:

- с 1967 г. остатки зависят от времени и имеет место существенное систематическое завышение слоя стока половодья, вычисленного по эмпирической зависимости (рис.1а);
- наклонная полоса рассеяния на рис.1б показывает, что отклонения от полученной эмпирической зависимости носят систематический характер: отрицательные остатки соответствуют большим по величине значениям расчетных слоев стока, положительные – малым, что свидетельствует о неточном определении свободного члена в уравнении;
- изгиб полосы рассеяния на рис.1в показывает, что в уравнении необходимо учесть нелинейность зависимости  $Y$  от  $X_1$ ;
- из рис.1г следует, что коэффициент перед  $X_2$  также определен неверно.

# Новая эмпирическая зависимость:

$$Y = 0.76X'_1 + 0.14 \cdot 10^{-5} X_2^2 + 13.8$$

где:  $X'_1 = X_{10} \cdot \gamma_p$  и

$X_{10}$  – средний максимальный снеговой запас в бассейне (мм), осредненный по 10 метеостанциям, для которых коэффициент корреляции снеговых запасов со стоком половодья  $>0.5$ ;

$\gamma_p = K_p \cdot \beta_p$  и

$K_p$  – модульный коэффициент приведенных запасов влаги в почве,

$\beta_p = \sin(\alpha' + 100)$ ,  $\text{tg } \alpha' = K_H$  и

$K_H$  – модульный коэффициент промерзания;

$X_2$  – слой стока за март (мм).

Коэффициент корреляции полученного уравнения равен 0.89