

Линейные регрессионные модели с гетероскедастичными и автокоррелированными остатками

1. Предпосылки метода наименьших квадратов
2. Гетероскедастичность, выявление и устранение
3. Автокорреляция, выявление и устранение
4. Мультиколлинеарность, выявление и устранение

Предпосылки метода наименьших квадратов

Условия Гаусса-Маркова

1. условие

$$M(\varepsilon_i) = 0$$

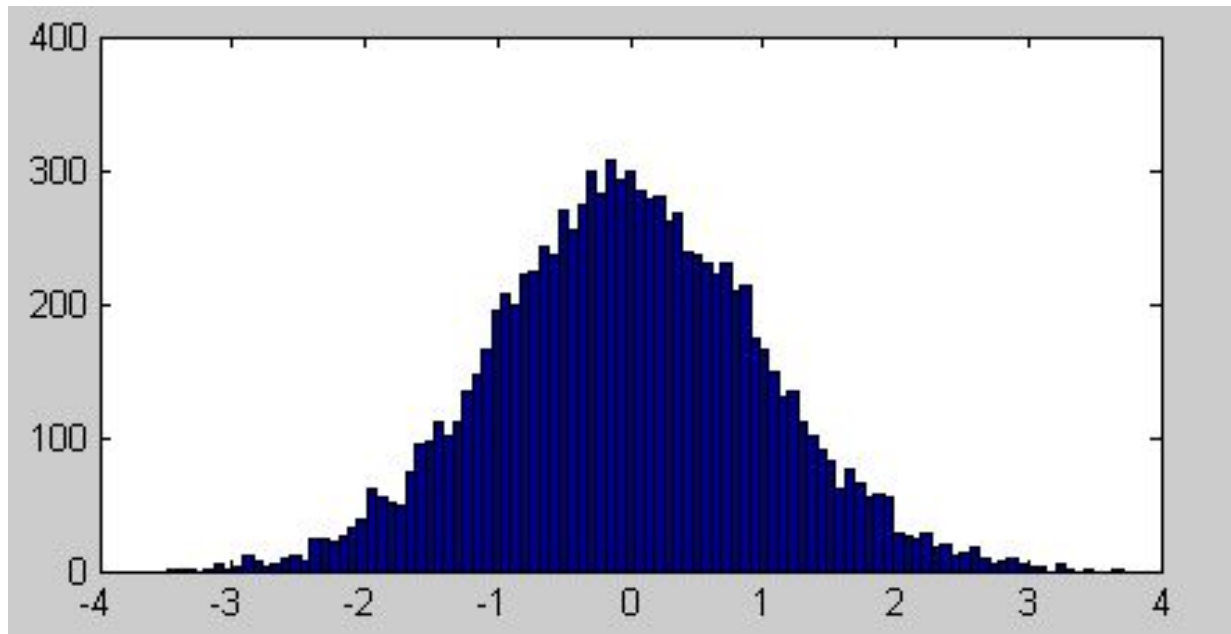
2. условие

$$D(\varepsilon_i) = \sigma^2$$

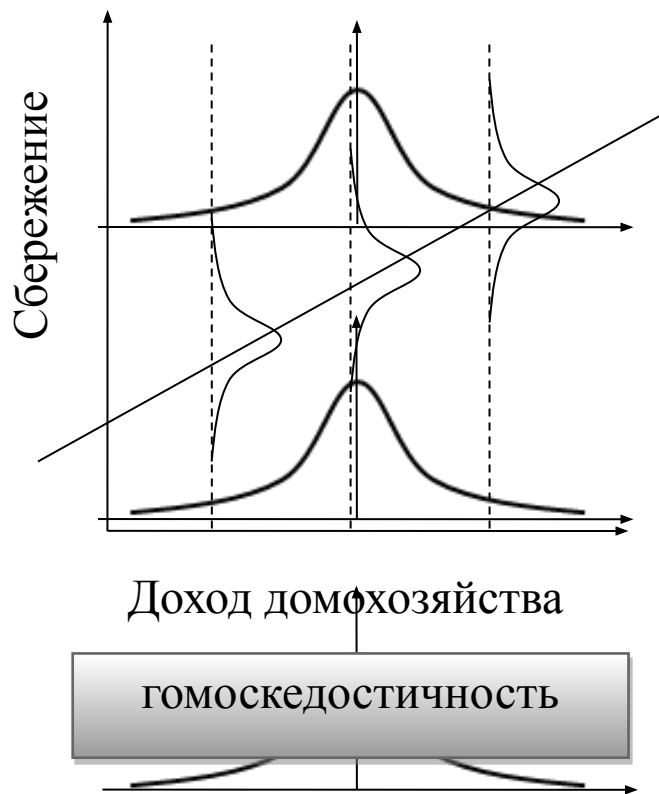
3. условие

$$M(\varepsilon_i \varepsilon_j) = 0 \ (i \neq j)$$

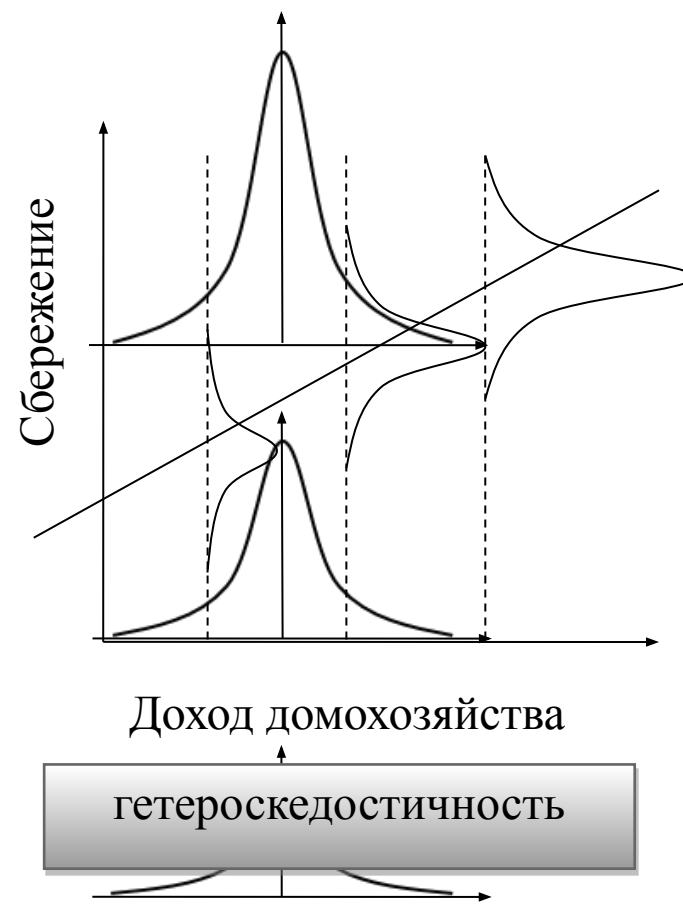
4. условие



Гетероскедстичность, выявление и устранение



гомоскедестичность

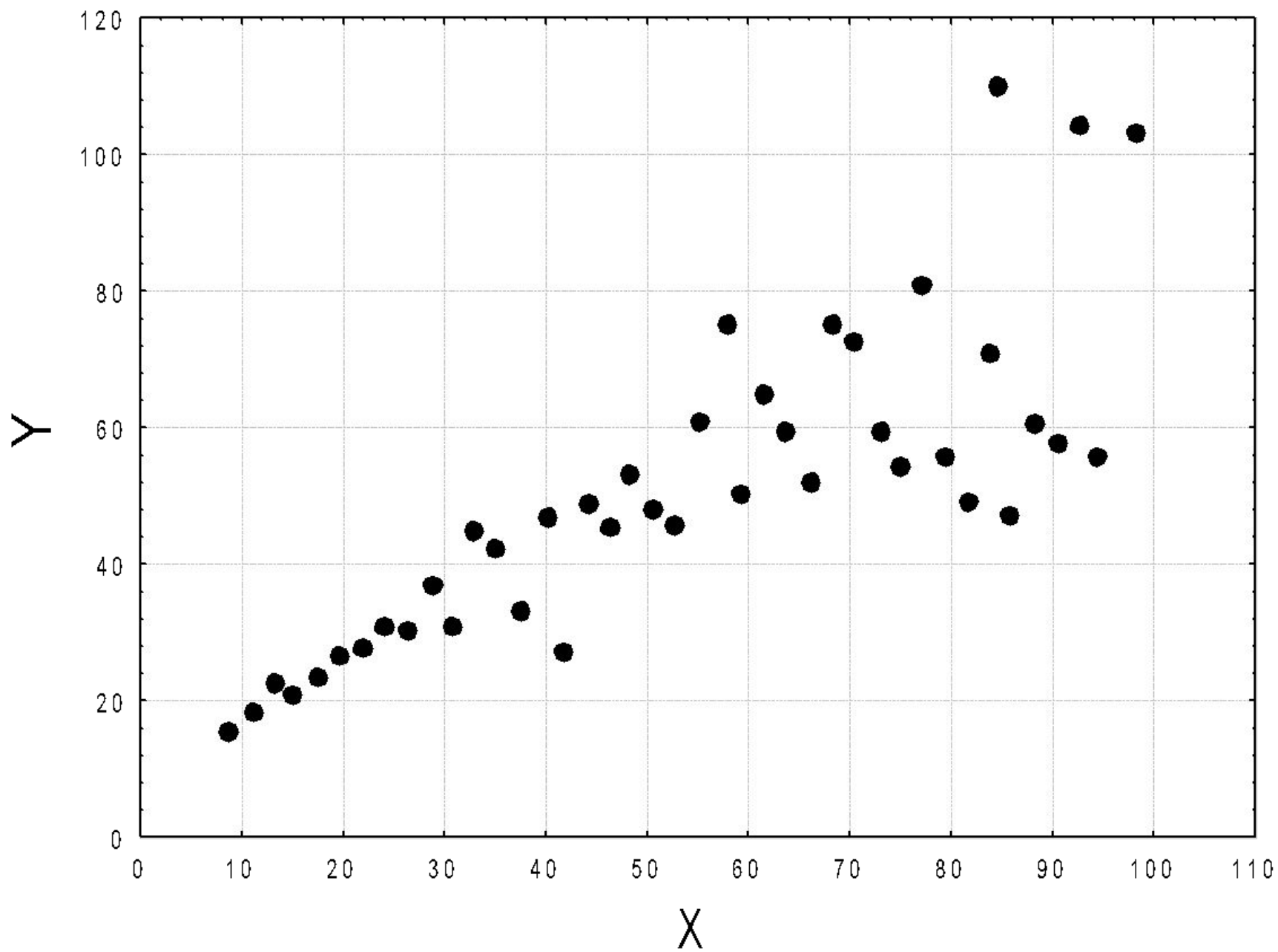


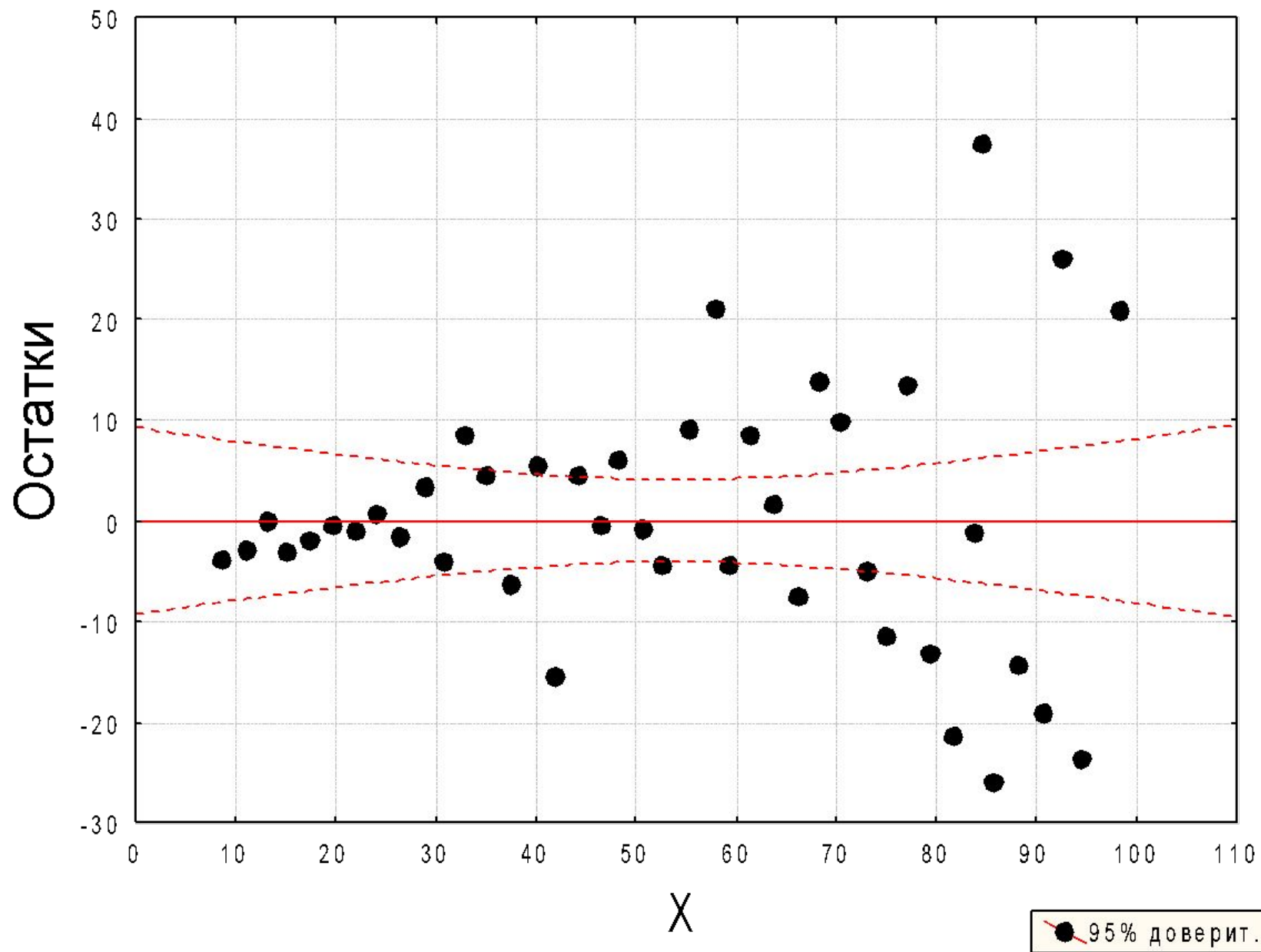
гетероскедестичность

Пример гетероскедстичности в пространственных данных: эффект масштаба

X – стоимость основных производственных фондов
(млн. руб.)

Y – прибыль предприятия (млн. руб.).



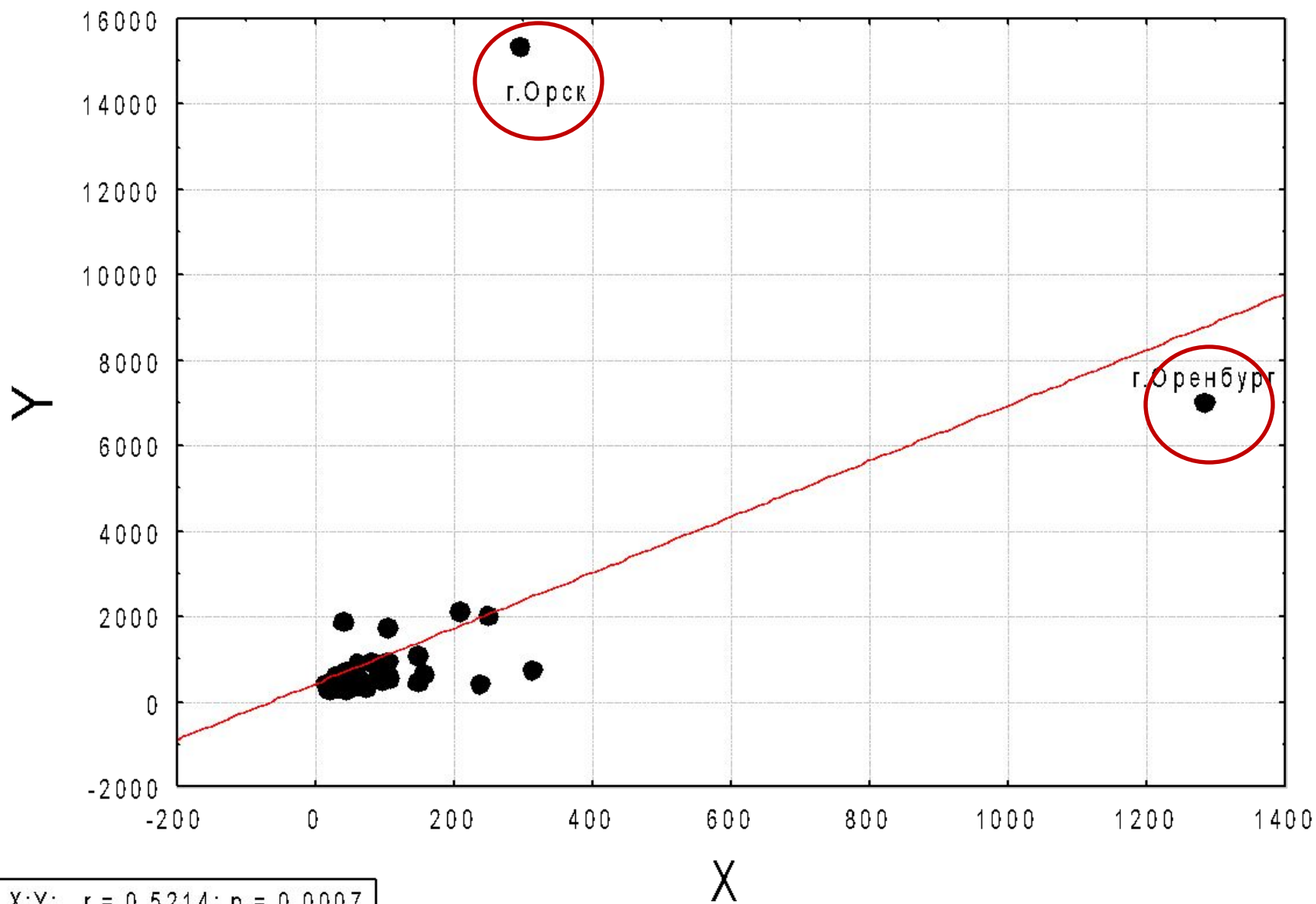


Пример гетероскедстичности в пространственных данных: эффект выбросов

X - численность официально зарегистрированных
безработных (чел.)

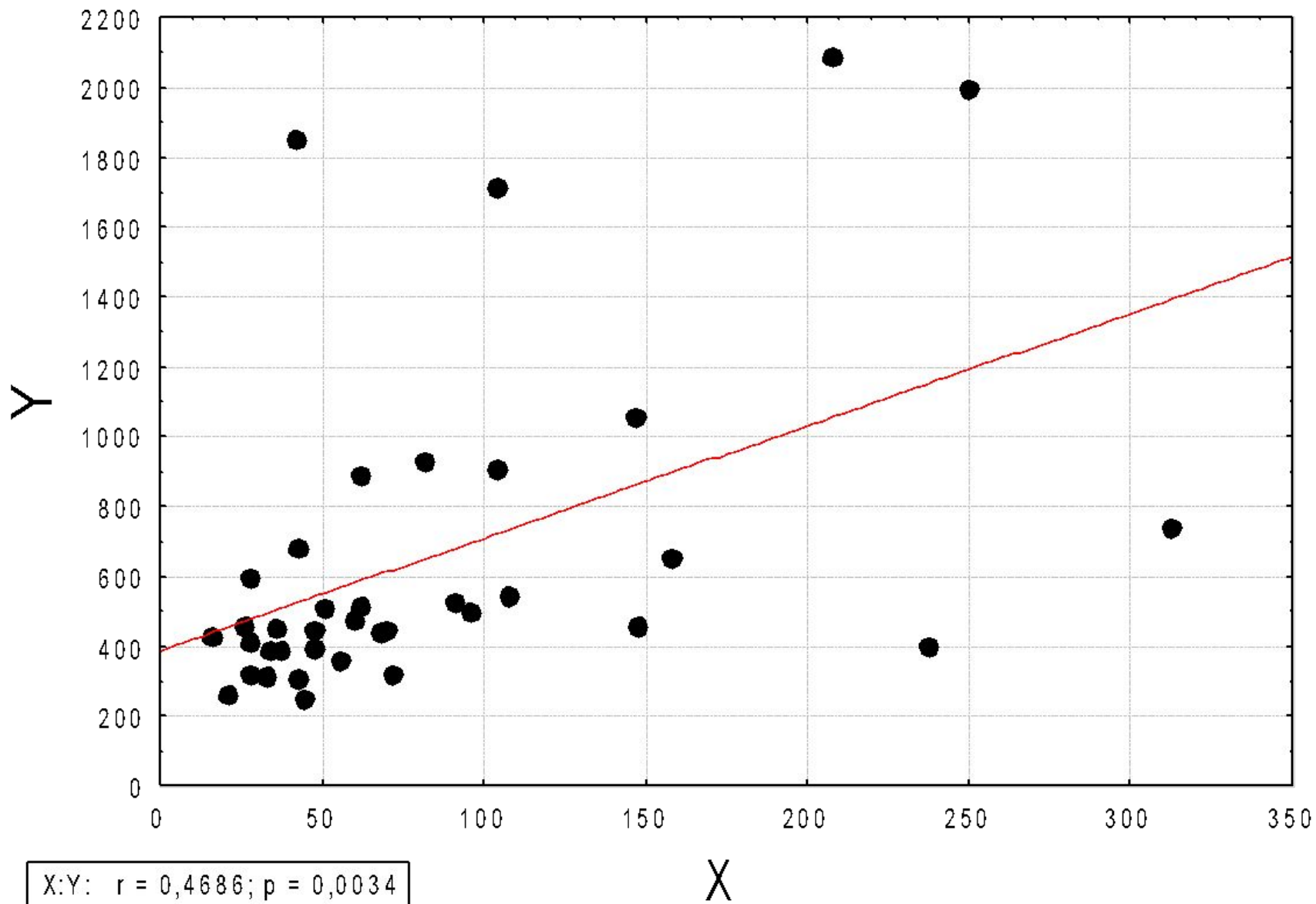
Y - число зарегистрированных преступлений (ед.).

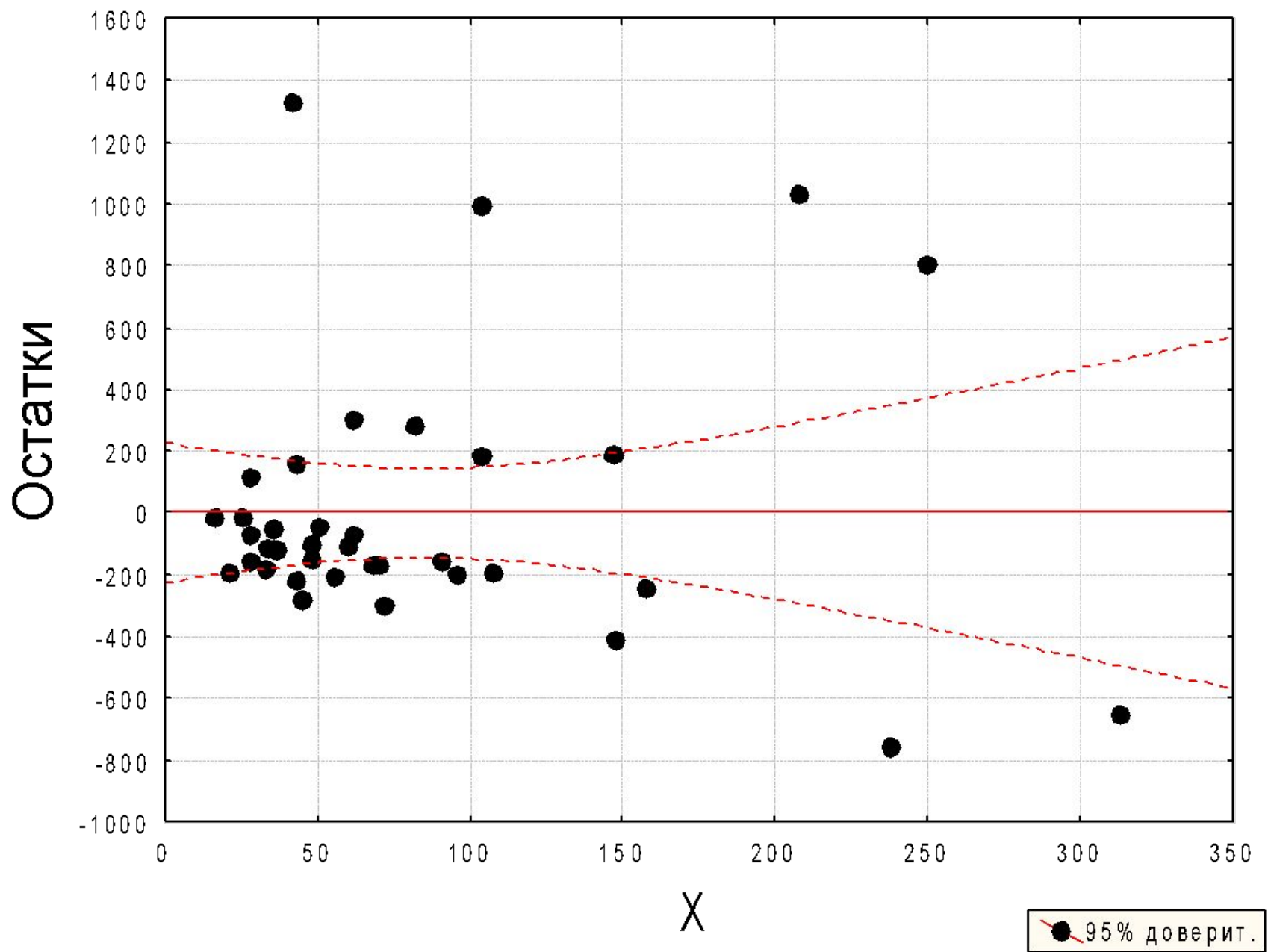
$$Y = 412,3346 + 6,5233 * X$$



X:Y: $r = 0,5214$; $p = 0,0007$

$$Y = 387,6437 + 3,2146 * x$$



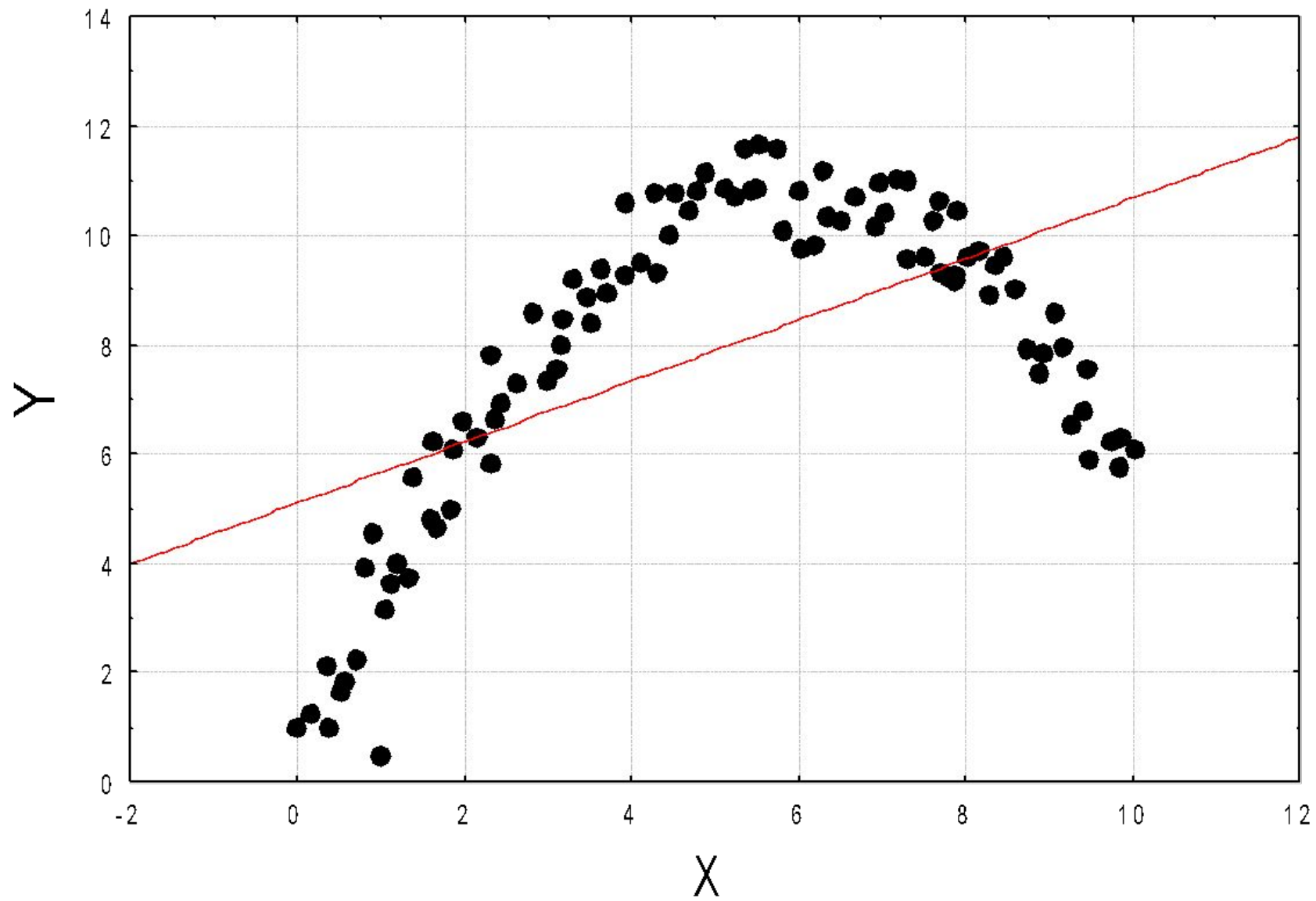


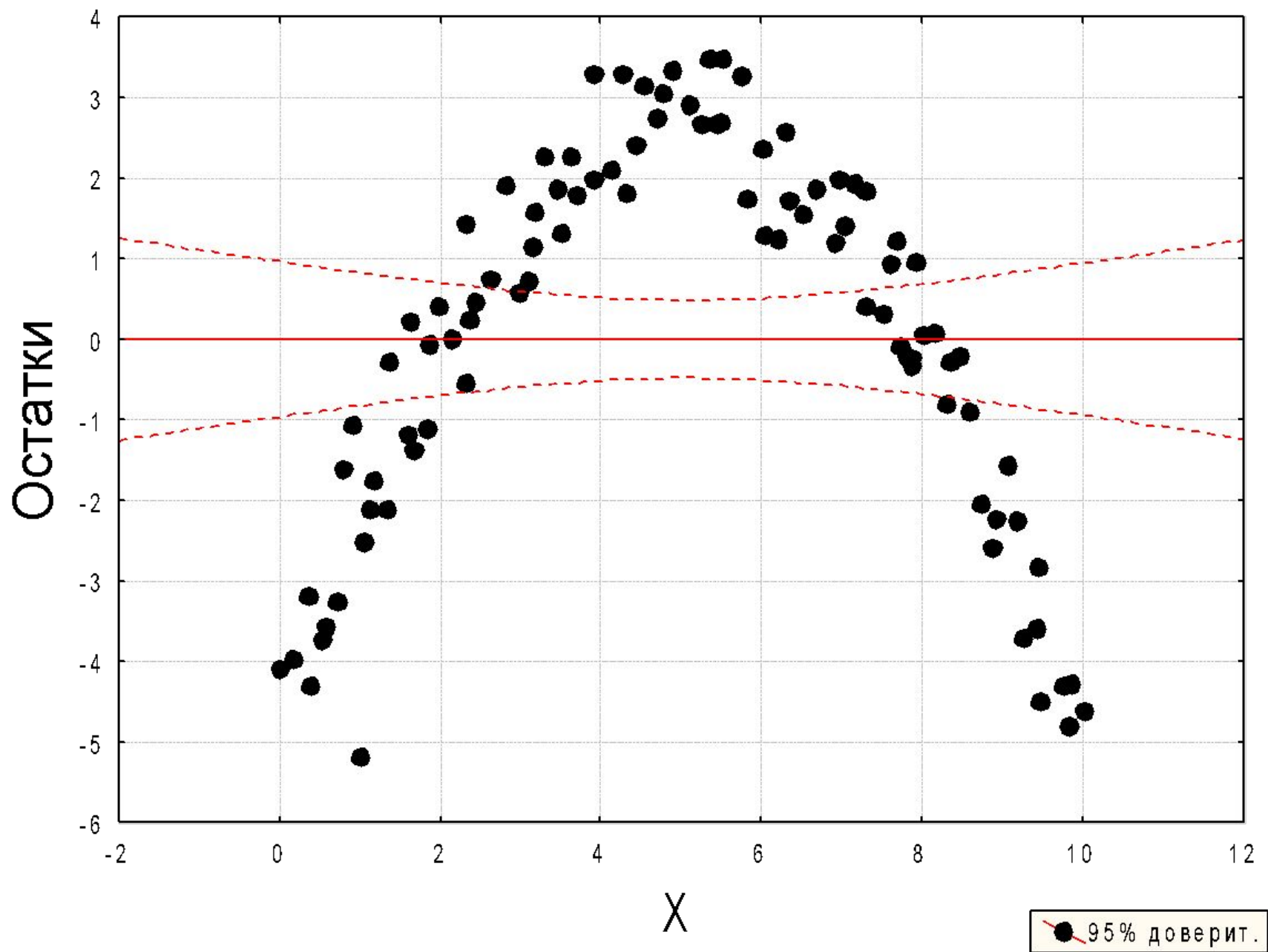
Пример гетероскедстичности в пространственных данных: неверная спецификация

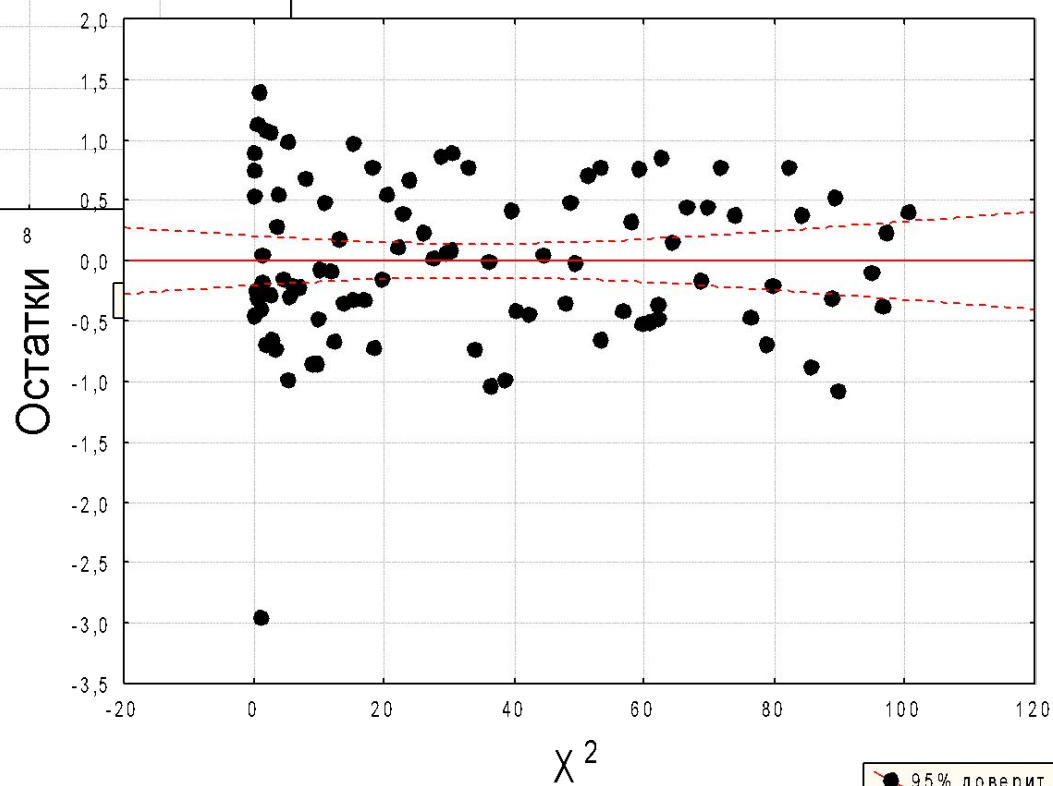
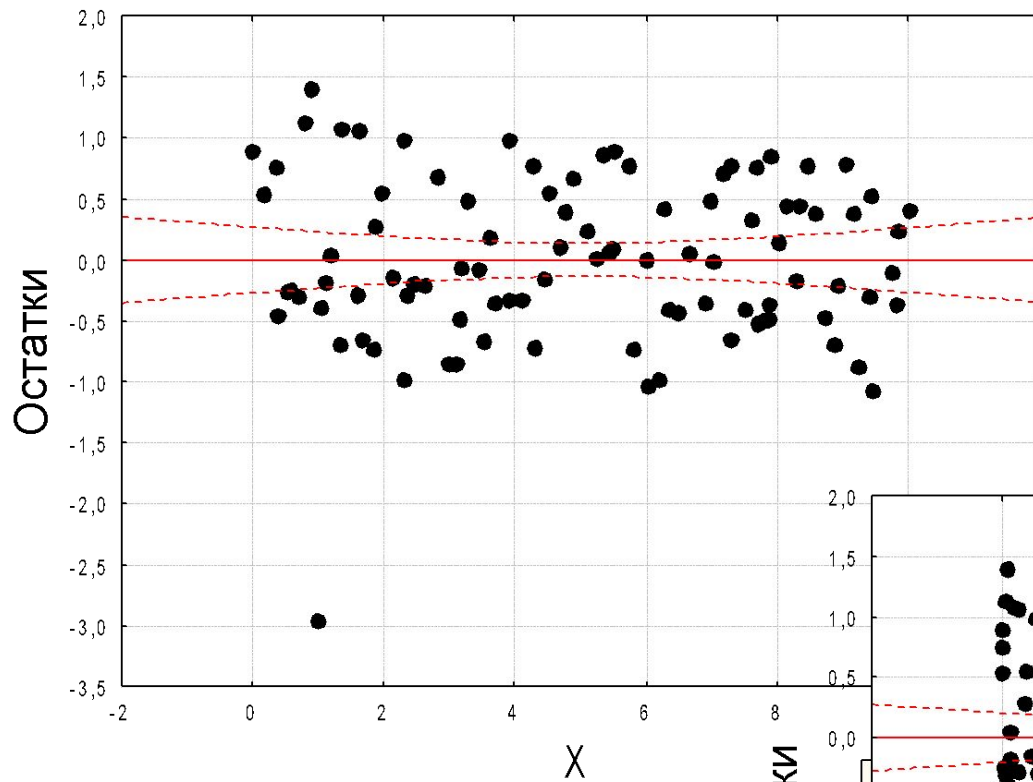
X – затраты на рекламу

Y – прибыль предприятия

$$Y = 5,1038 + 0,5573 * X$$



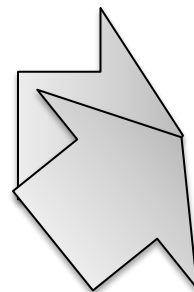
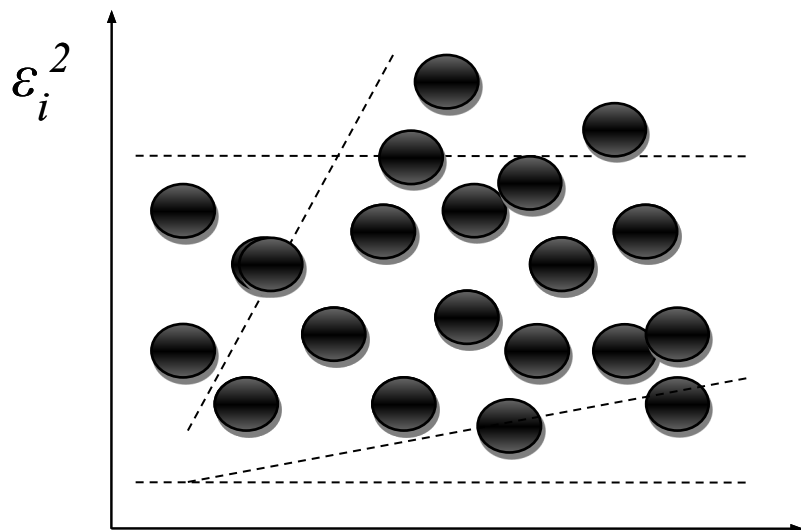




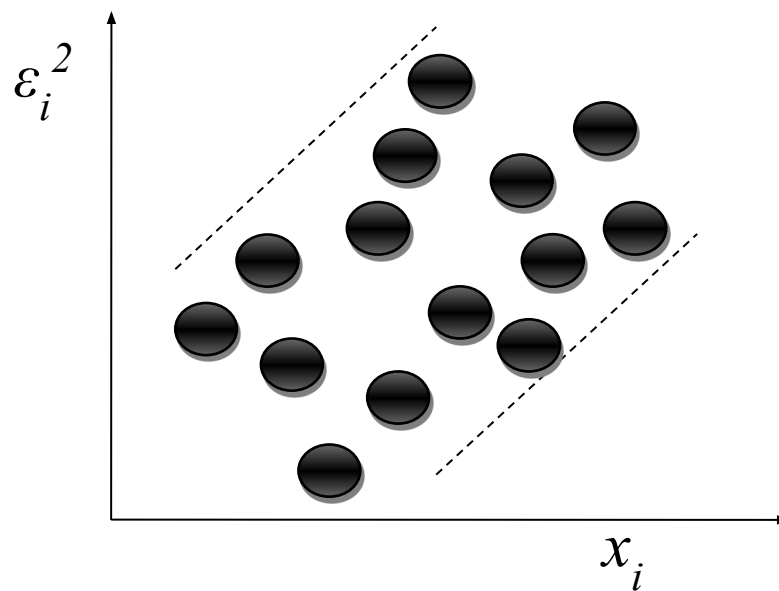
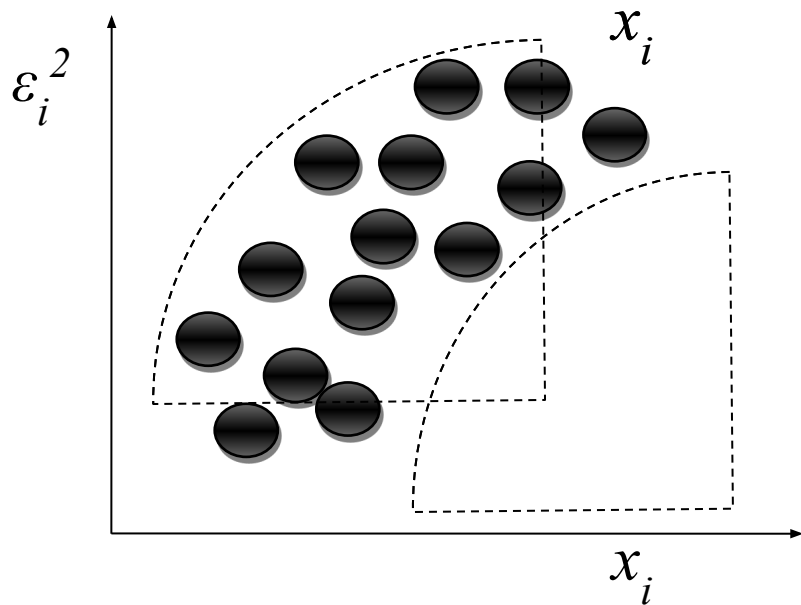
Методы обнаружения гетероскедастичности

- ✓ графический анализ отклонений
- ✓ тест ранговой корреляции Спирмена
- ✓ тест Парка
- ✓ тест Глейзера
- ✓ тест Голдфреда-Квандта
- ✓ тест Уайта

Графический анализ отклонений



гетероскедестичность



Тест ранговой корреляции Спирмена

1 этап. Значения x_i и ε_i ранжируются. Затем определяется коэффициент по формуле:

$$\rho = 1 - \frac{6 \times \sum_{i=1}^n (R_i^X - R_i^\varepsilon)^2}{n(n^2 - 1)}$$

R_i^X, R_i^ε - ранги по независимой переменной и случайным отклонениям;

n - объем изучаемой совокупности.

2 этап. Находится t -фактическое по формуле:

$$t_{\text{факт}} = \frac{\rho \sqrt{n-2}}{\sqrt{1-\rho^2}}$$

3 этап. $t_{\text{факт}}$ сравнивается $t_{\text{табл}}$ ($\alpha/2; \nu=n-2$). Если $t_{\text{факт}} > t_{\text{табл}}$ то необходимо отклонить гипотезу об отсутствии гетероскедестичности

Тест Парка

1 этап. Строится уравнение регрессии:

$$\tilde{y}_i = a_0 + a_1 x_i + \varepsilon_i$$

2 этап. Для каждого наблюдения определяется:

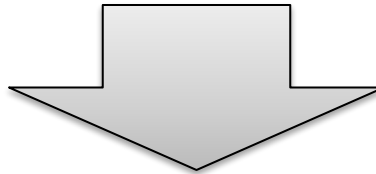
$$\ln \varepsilon_i^2 = \ln(y_i - \tilde{y}_i)^2$$

3 этап. Строится регрессионное уравнение:

$$\ln \varepsilon_i^2 = \alpha + \beta \ln x_i + u_i$$

4 этап. Проверяется статистическая значимость коэффициента β на основе t -статистики.

Если коэффициент β значим, то это означает наличие связи между $\ln \varepsilon_i^2$ и $\ln x_i$.



гетероскедостичности в статистических данных
присутствует

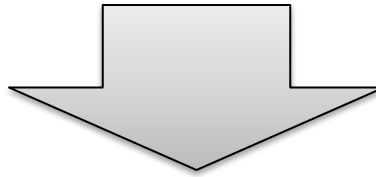
Тест Глейзера

1 этап. Строится уравнение: $\tilde{y}_i = a_0 + a_1 x_i + \varepsilon_i$

2 этап. Находят: $|\varepsilon_i| = |y_i - \tilde{y}_i|$

3 этап. Строится регрессия: $|\varepsilon_i| = \alpha + \beta x_i^k + u_i$

4 этап. С помощью t -критерия Стьюдента проверяют значимость коэффициента β , и если он значим $t_{\text{факт}} > t_{\text{табл}}$ то:



гетероскедстичности в статистических данных
присутствует

Тест Гольфельда-Квандта

- 1 этап.** Все n наблюдений упорядочиваются по величине X .
- 2 этап.** Вся упорядоченная выборка после этого разбивается на три подвыборки размером k , $(n-2k)$, k соответственно.
- 3 этап.** Оцениваются отдельные регрессии для первой подвыборки (k первых наблюдений) и для третьей подвыборки (k последних наблюдений).
- 4 этап.** Для сравнения соответствующих дисперсий строится F - стат

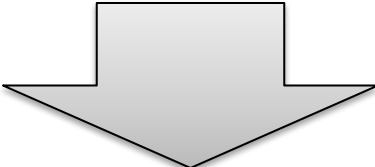
$$F_{\text{факт}} = \frac{\sum_{i=1}^k \varepsilon_i^2 / k - m - 1}{\sum_{i=n-k-1}^n \varepsilon_i^2 / k - m - 1} = \frac{\sum_{i=1}^k \varepsilon_i^2}{\sum_{i=n-k-1}^n \varepsilon_i^2}$$

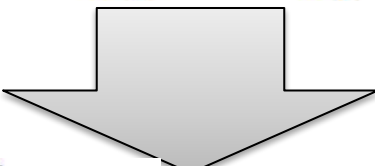
где: m - число объясняющих переменных в уравнении регрессии.

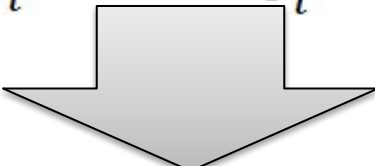
- 5 этап.** Если $F_{\text{факт}} > F_{\text{табл}}(\alpha, v_1=v_2=k-m-1)$, то гипотеза об отсутствии гетероскедстичности отклоняется.

Метод взвешенных наименьших квадратов

$$\tilde{y}_i = a_0 + a_1 x_i + \varepsilon_i \quad \text{разделим обе части уравнения на:} \quad \sigma_i = \sqrt{\sigma_i^2}$$


$$\frac{\tilde{y}_i}{\sigma_i} = a_0 \frac{1}{\sigma_i} + a_1 \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$$


$$\frac{\tilde{y}_i}{\sigma_i} = \tilde{y}_i^* \quad \frac{1}{\sigma_i} = z_i \quad \frac{x_i}{\sigma_i} = x_i^* \quad \frac{\varepsilon_i}{\sigma_i} = u_i^*$$

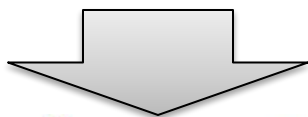

$$\tilde{y}_i^* = a_0 z_i + a_1 x_i^* + u_i$$

При этом для u_i выполняется условие гомоскедастичности.

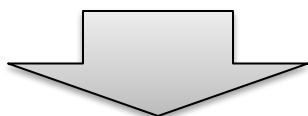
Дисперсии
пропорциональны x_i



$$\sqrt{x_i}$$

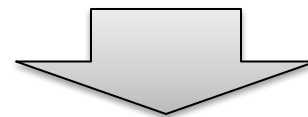


$$\frac{\tilde{y}_i}{\sqrt{x_i}} = a_0 \frac{1}{\sqrt{x_i}} + a_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}$$



$$\frac{\tilde{y}_i}{\sqrt{x_i}} = a_0 \frac{1}{\sqrt{x_i}} + a_1 \sqrt{x_i} + u_i$$

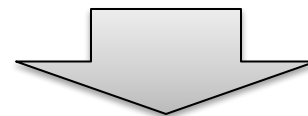
Дисперсии
пропорциональны x_i^2



$$x_i$$



$$\frac{\tilde{y}_i}{x_i} = a_0 \frac{1}{x_i} + a_1 + \frac{\varepsilon_i}{x_i}$$



$$\frac{\tilde{y}_i}{x_i} = a_0 \frac{1}{x_i} + a_1 + u_i$$

В ряде случаев для устранения гетероскедастичности необходимо изменить спецификацию модели (например, линейную на лог-линейную, мультипликативную на аддитивную и т. п.).